1-1-1985

The development and application of Latent Discriminant Analysis to marketing.

Narendra P. Mulani
University of Massachusetts Amherst

THE DEVELOPMENT AND APPLICATION OF LATENT DISCRIMINANT ANALYSIS TO MARKETING

A Dissertation Presented

By

NARENDRA MULANI

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 1985

School of Management

THE DEVELOPMENT AND APPLICATION OF LATENT DISCRIMINANT ANALYSIS TO MARKETING

A Dissertation Presented

by

Narendra Mulani

Approved as to Style and Content:

William R. Dillon, Chair, SOM

Thomas Madden, SOM

Anna Nagurney, SOM

Neil Firtle (Market Facts, Inc., New York)

D. Anthony Butterfield
Ph.D. Program Director, SOM

ACKNOWLEDGEMENTS

My thesis advisor, William R. Dillon, encouraged me to work on this very interesting modeling problem. For this and other types of support throughout my graduate years I owe much gratitude. Thanks are also due to Neil Firtle, Thomas Madden, and especially Anna Nagurney for their advice and encouragement. I owe a special debt to Dr. Nagurney for introducing me to the field of mathematical optimization.

My friends at the Management Research Center, Nancy Podolak and Louis Wigdor, gave me much support over the last four years. Rajiv Grover and Ajith Kumar kept me humble and were very helpful at the conceptual stage of this thesis. Dianne Peterson's expertise has made production of this thesis a much simpler task than anticipated.

At a personal level, Jehangir Talati and Jim Wilson have given me much. My parents' sacrifice and encouragement have made this possible. Finally, this thesis is dedicated to my wife Nita, who brought joy into my life.

Abstract

The Development and Application of Latent

Discriminant Analysis to Marketing

(September 1985)

Narendra Mulani, B.Com., Bombay University

M.B.A., Ph.D., University of Massachusetts

Directed by: Dr. William R. Dillon

The Latent Discriminant Model, a latent structure model that allows for overlapping groups, is formulated. Maximum Likelihood Estimates are derived and an algorithm for their solution is given. Numerical tests indicate that the algorithm adequately captures a given structure. Two empirical applications, perceptual mapping and constrained clustering, are illustrated in the Latent Discriminant framework.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

This thesis develops, tests and suggests applications of the Latent Discriminant Model. The Latent Discriminant Model falls under the general class of Latent Structure Models (Lazarsfeld and Henry, 1968). The model may also be studied as a particular finite mixture model (e.g., Everitt and Hand, 1981).

Most multivariate data of interest in marketing research has three modes: individuals rating objects on attributes. The Latent Discriminant model preserves the object model, a mode not normally preserved using standard analytic tools available to the researcher (i.e., factor analysis/principal components). Additionally the model allows one to mix categorical and continuous variables freely, within the same analysis.

The thesis is organized as follows. Chapter II contains a methodological review that shows the foundations from which the Latent Discriminant Model was formulated. Chapter III derives the Latent Discriminant model. Maximum likelihood equations are derived and an algorithm is suggested, along with some other results of interest. Chapter IV is devoted to a Monte Carlo simulation which tests whether the algorithm recovers a known structure. Chapter V suggests two applications of the model--perceptual mapping and constrained clustering. Implementation of these models in the Latent Discrimination framework is illustrated, and empirical examples are given. Finally Chapter VI summarizes the study.

CHAPTER II

METHODOLOGY

Introduction

This chapter begins by introducing latent structure analysis. The latent class model, goodness-of-fit criteria, estimation and identifiability are discussed in detail. The next section discusses Lazarsfeld's conceptualization of latent profile analysis. The chapter ends by discussing the limitations and capabilities of these models.

Latent Structure Analysis

Latent Structure Analysis was developed by Lazarsfeld (1959), and Lazarsfeld and Henry (1968). Although the theoretical development was complete, a general method of estimation was not available until recently and hence initial applications were few. Occasional papers did appear on the subject (McHugh 1956, Gibson 1959, McDonald 1962). These papers were primarily theoretical, delving into particular mathematical nuances with little focus on application. In the early 1970s an algorithmic breakthrough was achieved by Leo Goodman (1974). In a series of papers he clearly explicated how the latent class parameters could be estimated (Goodman 1974a,b). Since that time LSA has grown in popularity, especially in sociology. In marketing LSA was introduced by Green, Carmone and Wachpress (1976). In that paper, Green and his colleagues showed how the parameters of the latent class model could be estimated using CANDECOMP (Carroll and Chang 1970) and presented an empirical example. Other applied studies have followed: Nicosia (1977) discussed the Latent Class model and its use in marketing; Madden and Dillon (1982) used Latent Class analysis as a means for modeling causal relationships in categorical data; Dillon, Madden and Mulani (1983) tested various scaling models for categorical variables; and Grover and Dillon (1984) used Latent Class analysis as a framework for testing hypothesized hierarchical market structures. In large measure the growth in popularity of LSA in applied research settings is due to the work of Clogg (1977) who has made available a flexible program called Maximum Likelihood Latent Structure Analysis (MLLSA).

The Probabilistic Model

Assume that we have a situation that requires each respondent to generate k responses, and each response is associated with a single task. The data may be represented by a vector $\underline{u}$. Also assume that there are T true responses in the situation, where the vector for each true response may be represented as $\underset{\sim}{V}_t$. Note that a manifest vector $\underset{\sim}{u}_s$ does not necessarily correspond to one of the true vectors $\underset{\sim}{V}_t$. This discrepancy arises as a result of an "error" by one of the respondents. Using the notation P(•) for probabilities, the model can be expressed as

$$(1)\quad P(\underset{\sim}{u}_s) = \sum_{t=1}^{T} P(\underset{\sim}{u}_s | \underset{\sim}{V}_t) \cdot \Theta_t$$

where Θ_t is the probability that the t^{th} pattern vector occurs, i.e., the hypothetical population proportion of individuals that are in the t^{th} true response category with the restriction that

(2) $\sum_{t=1}^{T} \Theta_t = 1$

The conditional probabilities $P(\underset{\sim}{u}_s|\underset{\sim}{V}_t)$ are called the recruitment probabilities since they connect the manifest variables to the latent responses. These conditional probabilities can be represented in terms of response error probabilities as

(3) $P(\underset{\sim}{u}_s|\underset{\sim}{V}_t) = \prod_{k=1}^{K} \gamma_{kt}^{a_k}(1-\gamma_{kt})^{1-a_k}$

where γ_{kt} = the probability of the k^{th} response vector being in error for the t^{th} latent class vector,

$1-\gamma_{kt}$ = the probability of the k^{th} response being correct for the t^{th} latent class vector $(\gamma_{kt}+(1-\gamma_{kt})=1.0)$,

a_k = 1 if the k^{th} item in $\underset{\sim}{u}_s$ does not agree with the k^{th} item in $\underset{\sim}{V}_t$, that is, $u_{sk} \neq V_{tk}$, and

a_k = 0 if the k^{th} item in $\underset{\sim}{u}_s$ agrees with the k^{th} item in $\underset{\sim}{V}_t$, that is, $u_{sk}=V_{tk}$.

The probabilistic model states that a given manifest vector of responses $\underset{\sim}{u}_s$ is the weighted sum of the appropriate conditional probabilities given the true vector $\underset{\sim}{V}_t$ of responses within each class using weights Θ_t. Note the product form of (3): The implicit assumption in this form is that the k responses are locally independent given the latent class. Although this assumption may initially seem restrictive, it is a necessary condition for working with latent class models. In our previous definition of the probabilistic model (1), we have a weight Θ_t for each class. Assume the existence of a variable Θ with T levels

$(\Theta_1,\Theta_2,\ldots,\Theta_t)$. Local independence states that given the level of the latent variable Θ, the distribution of all the item responses are independent. This does not mean that the manifest item responses are unrelated to each other for the population of respondents. What it means is that the item responses for those respondents assigned to the t^{th} latent class are related to each other only through the t^{th} level of the latent variable Θ. A proof given by Lord and Novick (1968) is given for completeness.

Formally, local independence implies for responses $\underset{\sim}{u}_s=(u_1,u_2,,u_k)$ that for a given level Θ of the joint distribution Ω of $\underset{\sim}{u}_s$ is equal to the product of the marginal distributions γ; that is

$$(4)\quad \Omega(u_1,u_2,,u_k|\Theta) = \prod_{k=1}^{k} \gamma(u_k|\Theta)$$

Note the similarity between (4) and (3). Consider items 2,3,...,k from $\underset{\sim}{u}_s$ where

$$(5)\quad \Omega(u_2,u_3,\ldots,u_k|\Theta) = \prod_{k=2}^{k} \gamma_k(u_k|\Theta)$$

Dividing equation (5) into equation (4) yields the following result:

$$(6)\quad \gamma_1(u_1|\Theta;\ u_2,u_3,\ldots,u_k) = \gamma(u_1|\Theta)\ .$$

This shows that the conditional distribution γ_1 of u_1 for a given value of Θ, $u_1,u_2,\ldots,u_k$ does not depend on $u_2,u_3,\ldots,u_k$. This, of course, holds for each of the k items. To make an analogy with factor analysis, the assumption of local independence is equivalent to the factor analytic assumption that the factors explain the observed covariances across respondents. Alternatively we may state that the observed covariation between responses are due to a clustering of the population and the responses are independent within a cluster (Dillon and Goldstein 1984, p. 498).

The probabilistic definition given above can be translated to a set of accounting equations that allow parameterization and estimation of the model for an m-way cross classification. The main features are described below. A more detailed explanation can be found in Goodman (1974a).

We shall describe latent class analysis in the context of a cross classification of three manifest variables A, B and C. Assume the existence of latent variable X having T levels that explains the relationship among the manifest variables. Let $\hat{\Pi}_{ijk}$ denote the expected joint probability in the i,j,kth cell of the observed cross classification AxBxC. The latent class model represents the observed joint probability Π_{ijk} as a function of the latent class parameters Π_t^X, $\Pi_{it}^{\bar{A}X}$, $\Pi_{jt}^{\bar{B}X}$, $\Pi_{kt}^{\bar{C}X}$ where

Π_t^X is the probability that an individual will be a member of a given class t.

$\Pi_{it}^{\bar{A}X}$ is the conditional probability of observing the individual at level i of variable A given that the individual is in the t^{th} level of the latent variable X.

$\Pi_{jt}^{\bar{B}X}$ is the conditional probability of observing the individual at level j of variable B given that the individual is in the t^{th} level of the latent variable X.

$\Pi_{kt}^{\bar{C}X}$ is the conditional probability of observing the individual at level k of variable C given that the individual is in the t^{th} level of the latent variable X.

In short, Π_t^X gives the proportion of individuals in latent class t. $\Pi_{it}^{\bar{A}X}$, $\Pi_{jt}^{\bar{B}X}$, and $\Pi_{kt}^{\bar{C}X}$ give the distribution of the i^{th} level of A, the j^{th} level of B, and the k^{th} level of C within the t^{th} latent class.

The observed joint probability Π_{ijk} can be recovered from the latent parameters by summing over the T latent classes for the unobservable joint probabilities

(7) $$\Pi_{ijk} = \sum_{t=1}^{T} \Pi_{ijkt}^{ABCX}$$

Using standard laws of probability we can write the unobservable joint probability Π_{ijkt}^{ABCX} in terms of the latent class proportion Π_t^X and the conditionals $\Pi_{it}^{\bar{A}X}$, $\Pi_{jt}^{\bar{B}X}$, $\Pi_{kt}^{\bar{C}X}$ as follows.

(8) $$\Pi_{ijk} = \sum_{t=1}^{T} \Pi_{ijkt}^{ABCX} = \sum_{t=1}^{T} \Pi_t^X \Pi_{it}^{\bar{A}X} \Pi_{jt}^{\bar{B}X} \Pi_{kt}^{\bar{C}X}$$

Equation (8) is referred to as the fundamental equation of latent structure analysis by Lazarsfeld and Henry (1968). As the latent class parameters are probabilities they are subject to some restrictions, specifically

(9a) $$\sum_{t=1}^{T} \Pi_t^X = 1$$

(9b) $$\sum_{i=1}^{I} \Pi_{it}^{\bar{A}X} = 1.$$

(9c) $\sum_{j=1}^{J} \Pi_{jt}^{\bar{B}X} = 1.$

(9d) $\sum_{k=1}^{K} \Pi_{kt}^{\bar{C}X} = 1.$

(10a) $\sum_{t=1}^{T} \Pi_{t}^{X}\Pi_{it}^{\bar{A}X} = A_i$

(10b) $\sum_{t=1}^{T} \Pi_{t}^{X}\Pi_{jt}^{\bar{B}X} = B_j$

(10c) $\sum_{t=1}^{T} \Pi_{t}^{X}\Pi_{kt}^{\bar{C}X} = C_k$

Equation (10) states that the sum of the product of the latent class proportions and conditional probabilities of a given manifest variable results in the observed marginal for that variable.

Let $\Pi_{ijkt}^{ABC\bar{X}}$ denote the conditional probability that X is at level t given that A, B, and C are at level i,j,k. Then by definition

(11) $\Pi_{ijkt}^{ABC\bar{X}} = [\Pi_{ijkt}^{ABCX}/\Pi_{ijk}]$

From the relationship shown in equation (11) we can write Π_{t}^{X} as

(12) $\Pi_{t}^{X} = \sum_{i}\sum_{j}\sum_{k} \Pi_{ijk}\Pi_{ijkt}^{ABC\bar{X}}$

Similarly we can write the conditional probabilities as:

(13) $\Pi_{it}^{\bar{A}X} = (1/\Pi_{t}^{X})\cdot\sum_{j}\sum_{k} \Pi_{ijk}\Pi_{ijkt}^{ABC\bar{X}}$

(14) $\Pi_{jt}^{\bar{B}X} = (1/\Pi_{t}^{X})\cdot\sum_{i}\sum_{k} \Pi_{ijk}\Pi_{ijkt}^{ABC\bar{X}}$

$$(15)\quad \pi_{kt}^{\bar{C}X} = (1/\pi_t^X)\cdot\sum_i\sum_j \pi_{ijk}\pi_{ijkt}^{ABC\bar{X}}$$

Estimation

The above system of equations (7)-(15) suggests an iterative proportional scaling algorithm (Goodman 1974a,b) that produces maximum likelihood estimates. It turns out that this algorithm is a special case of the EM algorithm (Dempster, Laird and Rubin 1977) that yields maximum likelihood estimates for incomplete data. In the present case the missing data are the π_t^X, the unobserved proportion of individuals in each class. The EM algorithm is a simple and powerful method to estimate incomplete data. The system recognizes that if the missing data were known, the sufficient statistics would be used for maximum likelihood estimation of the parameters. On the other hand, if the parameters of the model were known, the missing data can be estimated via the conditional expectations given the observed data. We therefore have a two step procedure. In the E step, the missing data are estimated given the current estimates of the parameters of the model. In the M step, the maximum likelihood estimates are obtained given the expectations of the missing data.

This suggests a simple algorithm in our case. Start with an initial value for $\hat{\underset{\sim}{\Pi}}$, where $\hat{\underset{\sim}{\Pi}} = (\hat{\pi}_t^X, \hat{\pi}_{it}^{\bar{A}X}, \hat{\pi}_{jt}^{\bar{B}X}, \hat{\pi}_{kt}^{\bar{C}X})$ are the maximum likelihood estimates of our parameters. Using equation (8) we have

$$(16)\quad \hat{\pi}_{ijkt}^{ABCX} = \hat{\pi}_t^X \hat{\pi}_{it}^{\bar{A}X} \hat{\pi}_{jt}^{\bar{B}X} \hat{\pi}_{kt}^{\bar{C}X}$$

then

(17) $\hat{\Pi}_{ijk} = \sum_{t=1}^{T} \hat{\Pi}^{ABCX}_{ijkt}$

From equation (11) we note

(18) $\hat{\Pi}^{ABC\bar{X}}_{ijkt} = (\hat{\Pi}^{ABCX}_{ijkt}/\hat{\Pi}_{ijk})$

Similarly, using (12)-(15) we obtain the following

(19) $\hat{\Pi}^{X}_{t} = \sum_{i}\sum_{j}\sum_{k} \hat{\Pi}_{ijk}\hat{\Pi}^{ABC\bar{X}}_{ijkt}$

(20) $\hat{\Pi}^{\bar{A}X}_{it} = (1/\hat{\Pi}^{X}_{t})\cdot \sum_{j}\sum_{k} \Pi_{ijk}\hat{\Pi}^{ABC\bar{X}}_{ijkt}$

(21) $\hat{\Pi}^{\bar{B}X}_{jt} = (1/\hat{\Pi}^{X}_{t})\cdot \sum_{i}\sum_{k} \Pi_{ijk}\hat{\Pi}^{ABC\bar{X}}_{ijkt}$

(22) $\hat{\Pi}^{\bar{C}X}_{kt} = (1/\hat{\Pi}^{X}_{t})\cdot \sum_{i}\sum_{j} \Pi_{ijk}\hat{\Pi}^{ABC\bar{X}}_{ijkt}$

We use the results from (19)-(22) to reestimate (16)-(18) until the parameters converge to some stable estimate. This solution will either be a maximum likelihood solution, or some terminal solution, in which one or more of the parameters have been estimated at 0 or 1. Haberman (1977) analyzes the above system as a nonlinear programming problem and shows that a true maximum may be reached for large sample problems under certain conditions.

Identification

In general, a latent class model will be identified if the number of non-redundant cells is greater than the number of non-redundant parameters. Because of the constraints mentioned in (9) and (10), we need only consider T-1 of the Π_t^X, I-1 of the $\Pi_{it}^{\bar{A}X}$, etc. Thus for the AxBxC cross-classification at I,J,K levels, we need only consider

(23) $T-1 + (I+J+K-3)T = (I+J+K-2)T-1$

parameters. The necessary condition for model identification is:

(24) $IJK-1 \geq (I+J+K-2)T-1$

McHugh (1956) and Goodman (1974a) suggest a sufficient test for local identifiability that should be undertaken when the inequality shown in (24) is met. The model will be locally identifable if, for nonredundant parameters $\Theta \varepsilon \underset{\sim}{\Pi}$ the Jacobian matrix

(25) $[\partial\Pi_{ijk}/\partial\Theta]$

is of full rank. If the basic set (i.e., the set of nonredundant parameters) spans the latent space, we will get a full rank result. Lower ranks signify a misspecified latent model.

Goodness-of-fit

Goodness-of-fit for the models are tested by the likelihood ratio tests (L^2) and the Pearson χ^2 statistics

(26) $L^2 = 2n\sum p_{ijk}\ell n(p_{ijk}/\hat{\Pi}_{ijk})$

(27) $X^2 = n\sum(p_{ijk}-\hat{\Pi}_{ijk})^2/\hat{\Pi}_{ijk}$

where p_{ijk} is the observed proportion in cell ijk and $\hat{\Pi}_{ijk}$ is the maximum likelihood estimate of Π_{ijk}. Both the L^2 and χ^2 statistics are based on large sample theory. For very large samples, it has been shown that the likelihood-ratio statistic is preferred to the Pearson statistic (Haberman 1978). However, it is unclear which is to be preferred in samples that are not large. The recommended practice is to use both statistics. If they lead to different conclusions, then the large sample approximation may not be justified.

Both test statistics described above are a direct function of sample size. With moderate-to-large sample sizes, virtually any restricted model would be rejected as statistically untenable in favor of a less restricted model, even though the discrepancies between the predicted cell frequencies under the competing model are trivial. Consequently, measures of fit not directly based on the magnitude of the L^2 or χ^2 statistics can prove informative. Several useful indices of fit are:

$F_i = L^2(M_i)$. A quantity analogous to the F-statistic used in regression analysis (Haberman 1978).

$R^2 = [L^2(M_o)-L^2(M_i)]/L^2(M_o)$. A measure of fit which simply reflects the percent improvement in formal goodness-of-fit of a less restricted model M_i over the restricted model M_o of independence. R^2 is bounded by zero and unity (Goodman 1971, 1972).

$R^2_\omega = [F_o-F_i]/F_o$. A measure ф ffit which reflects goodness-of-fit as well as parsimony. This measure has an upper bound of 1.0, but in contrast to the R^2 measure, it can decrease in value when restrictions are lifted from the model if the improvement in goodness-of-fit is not commensurate with the loss of degrees of freedom (Bonnett and Bentler 1980).

$\Delta = \sum \Pi^{\overline{ABCX}}_{ijkt} P_{ijk}$. This measure gives the percentage of the sample correctly allocated into the classes of X. $\Pi^{\overline{ABCX}}_{ijkt}$ is the modal probability and P_{ijk} is the observed proportion corresponding to cell ijk and the summation is over all cells (Clogg 1977).

$\lambda = (E_1 - E_2)/E_1$. This measure is a symmetric index of association which gives the proportional reduction-in-error of the fitted model (Goodman and Kruskal 1954). In the context of the latent class model, E_2 is defined as the total expected error rate of assignment obtained from the use of the modal probabilities $\Pi_{ijkt}^{\overline{ABCX}}$, $E_2 = (1 - \Pi_{ijkt}^{\overline{ABCX}})$; E_1 is the expected proportion of errors obtained from the use of the model unconditional probabilities Π_{t*}^{X}--i.e., $E_1 = (1 - \Pi_{t*}^{X})$ (Clogg 1977).

The R^2 and R_ω^2 measures are incremental fix indices, which are particularly well suited for use in exploratory stages of model fitting, whereas the Δ and λ measures reflect classification indices, which are particularly well suited in judging the quality of the latent classes. Since the null model of independence M_0 is equivalent to a T=1 class model, the indexes R^2 and R_ω^2 represent indexes of increment in fit obtained by using t=2,3,...,T latent classes rather than one. Thus, we can view model M_0 as the completely restricted latent class model and evaluate alternative unrestricted models, which vary in the number of hypothesized latent classes, on the basis of their R^2 and R_ω^2 indexes.

Restricted Models

Latent structure analysis may also be used to confirm hypotheses. Various restricted models may be tested through LSA. Clogg's (1977) program allows for three basic types of parameter restrictions:
(1) Fixed parameters, which are known and not estimated. (2) Constrained parameters, which are not known, but assumed equal to some other

parameters. (3) Free parameters, which are unconstrained. An excellent discussion of the use and interpretation of restricted models is given in Clogg (1979). This completes the description of the latent class model. We now turn to Latent Profile Analysis.

Latent Profile Analysis

The Latent Structure model discussed in the previous sections can be used when the data consist of categorical manifest variables. It is common to face the problem of classifying a sample where the manifest variables are continuous. Common examples of such continuous variables are income, age and IQ. Gibson (1959), Lazarsfeld and Henry (1968), McDonald (1962), and Anderson (1982) show how the latent class equations may be used to explain the variation between continuous variables by classifying each sample vector into discrete subpopulations within which they are independent. In factor analysis we study the means of the variables $\mu=E(Y_i)$ and the covariances $\sigma_{ij}=E(Y_iY_j)-E(Y_i)E(Y_j)$, under the assumption that the joint distribution of the manifest variables is multivariate normal. Latent Profile Analysis explicitly recognizes the fact that the manifest variables Y_i are not normally distributed across the population. Rather, the manifest variables follow some specific distribution <u>within</u> each of the subpopulations (latent classes).

Assume that there are T classes in the population, t=1,2,...,T. Let μ_i^t denote the mean for the i^{th} variable in the t^{th} latent class. Similarly, let σ_i^t denote the within class variances for the i^{th} variable in the t^{th} latent class (the model assumes that the within class covariances are zero by the axiom of local independence). The general form of the model is:

$$(28)\ F(\underset{\sim}{Y}) = \sum_{t=1}^{T} P(X_t)\cdot \prod_{n=1}^{N} F_{it}(Y_i)$$

where $\underset{\sim}{Y}$ is the observed profile of scores on N attributes, $P(X_t)$ is the proportion of individuals in the t^{th} latent class, and $F_{it}(Y_i)$ is the density function of attribute i in latent class t.

The overall mean μ_i is the weighted sum of the within class means

$$(29)\ E(Y_i) = \mu_i = \sum_{t=1}^{T} P(X_t)\cdot\mu_i^t$$

and the variance

$$(30)\ \sigma_i^2 = E(Y_i-\mu_i)^2 = \sum_{t=1}^{T} P(X_t)E(Y_i-\mu_{it})^2 + \sum_{t=1}^{T} P(X_t)(\mu_{it}-\mu_i)^2$$

Recruitment of Individuals in the Latent Profile Model

The latent profile model theoretically partitions the population into classes, and if we know the within class distributions of the variables, it is straightforward to classify individuals. Following Lazarsfeld and Henry (1968), let $F_{it}(Y_i)$ be the probability density function of (Y_i) within class t. Thus

$$(31)\ \int_{-\infty}^{Z} F_{it}(Y_i)dy$$

is the probability that $Y_i \leq Z$ for those individuals that belong to class t. For a vector $(\underset{\sim}{Y})$, the manifest density function would be:

$$(32)\quad F(\underset{\sim}{Y}) = \sum_{t=1}^{T} P(X_t)F_t(\underset{\sim}{Y})$$

We classify an individual into the class for which the above density is a maximum. The probability that a respondent belongs to class t is:

$$(33)\quad P(X_t|\underset{\sim}{Y}) = (P(X_t)F_t(\underset{\sim}{Y})/F(\underset{\sim}{Y}))\forall_{t\varepsilon T}$$

The most likely class is the class "t" for which equation (33) is a maximum. Gibson (1959) and Lazarsfeld and Henry (1968) use complex matrix manipulations to estimate the parameters of the latent profile model. Simpler methods can be derived using the EM algorithm. Wolfe (1970) describes a simple estimation procedure based on the assumption that the within class distributions are normal.

In this chapter, we have reviewed existing variants of the latent structure model, namely the latent class and latent profile model. The latent class model can analyze data in the form of a multidimensional contingency table, i.e., the manifest variables are all assumed to be categorical. The latent profile model assumes that the input data are in the form of continuous variables that are distributed normally <u>within</u> each class. Both models cannot handle a situation where we have categorical as well as continuous manifest variables, a situation that occurs often in practice. Local independence is an axiom of the latent structure framework and theoretically there is no reason why we should not envision a latent structure model that allows for both categorical as well as continuous variables in the same model framework. The only constraint on the specification of the distributions of the manifest variables is that they follow some distribution from the exponential family within each of the latent classes. This constraint is purely pragmatic, as maximum likelihood estimation of the parameters of a model is possible via the EM algorithm as long as the variables follow the within class distributional assumptions stated above (Sundberg 1974).

The present framework of latent class analysis would be more attractive to marketing researchers if there was provision for a grouping variable. Clogg and Goodman (1984) discuss simultaneous latent class analysis in several groups. They do not allow groups to overlap, and primarily use their model as a means for testing various scaling models across the groups. By specifying a general model that allows for group overlap at the discretion of the analyst, such a model could potentially be used for various marketing problems that occur in the group context such as product choice, perceptual mapping and segmentation analysis. The major thesis of this dissertation is to propose a model that allows for a grouping variable and a combination of categorical and continuous manifest variables in the same model. This model, which we choose to call that Latent Discriminant Model is discussed in the next chapter.

CHAPTER III

THE LATENT DISCRIMINANT MODEL

Introduction

In this chapter we introduce and formulate the Latent Discriminant Model. The general model is defined, Maximum Likelihood equations are derived, and an algorithm for estimation is presented. The chapter ends with a discussion of goodness-of-fit, identification conditions, and some other useful results.

The Latent Discriminant Model

The latent discriminant model attempts to combine latent structure analysis and discriminant analysis into a unifying framework. In combining these two methodologies the proposed model can be used in application settings that could not be satisfactorily handled by either latent structure or discriminant analysis.

The Model

Let $\underset{\sim}{Y}$ be a response vector of order p, where the individual Y_i variables can be polytomous or continuous. Assume that each of N individuals belong to one of K groups, and devote the grouping variable by G_k, k=1,2,...,k. Next, introduce a latent variable X consisting of T classes. The variable is latent in that the classification of each response with respect to this variable is not directly observable. Thus, the observed data is of the form $[\underset{\sim}{G}|\underset{\sim}{Y}]$, where $\underset{\sim}{G}$ is an (Nx1) indicator of group membership ($N=n_1+n_2+...+n_k$) and $\underset{\sim}{Y}$ is an (Nxp) matrix of measurements on the N individuals. Given the individual group membership, G_k, a simple explanation for the relationship between the p variables in $\underset{\sim}{Y}$ is found by associating each response vector with one of the T latent classes of the variable X. Specifically the Latent Discriminant Model parameterizes the probability of observing a particular realization of the attribute vector, $\underset{\sim}{y}$, given the individual group membership, G_k, as follows

$$(34) \quad P(\underset{\sim}{y}|G_k) = \sum_{t=1}^{T} \prod_{i=1}^{P} P_i(y_i|\Theta_{it}) \cdot P(X_t|G_k)$$

where $P_i(y_i|\Theta_{it})$ is the marginal density of variable y_i. Note that $P_i(y_i|\Theta_{it})$ is the marginal density of y_i which depends on the latent parameter vector Θ_{it}, but not on the individual's group membership, G_k. The $P(X_t|G_k)$ are the mixing proportions, which depend on which G_k is being considered. Thus the density of an observed vector $\underset{\sim}{Y}$ within each latent class is estimated in equation (34) as the product of the conditional densities of the p measures within each latent class. Restrictions similar to those used in latent class model representation are also in place, for example:

$$(35) \quad \sum_{t=1}^{T} P(X_t|G_k) = 1,$$

which states that the probability of observing class t given a group k should sum to unity over all classes within each group; and

(36) $\sum_{m=1}^{M} P(y_{imt}) = 1,$

which states that for any polytomous variable i having m levels, the conditional probability of observing y given a latent class should sum to unity.

Derivation of Maximum Likelihood Estimates

We now derive the equations for the maximum likelihood estimates for the parameters of the latent discriminant model. For brevity and ease of notation, let $g(\underset{\sim}{y};\underset{\sim}{\theta}_t)$ replace $p_i(y_i|\Theta_{it})$ and $f(y_{jk})$ replace $P(\underset{\sim}{y}|G_k)$ in (34). We can now rewrite (34) as:

(37) $f(y_{jk}) = \sum_{t=1}^{T} P(X_t|G_k)g_t(\underset{\sim}{y}_{jk};\underset{\sim}{\theta}_t)$

The Likelihood for the sample is:

(38) $L(\underset{\sim}{y}|G_k;\ X_t|G_k,\ \underset{\sim}{\theta}_1,\ldots,\underset{\sim}{\theta}_T) = \prod_{k=1}^{k}\ \prod_{j=1}^{n_k} [\sum_{t=1}^{T} P(X_t|G_k)g_t(\underset{\sim}{y}_{jk};\underset{\sim}{\theta}_t)]$

Taking logarithms of the likelihood function and adding Lagrangian constraints on the parameters $P(X_t|G_k)$, we get:

(39) $L(\underset{\sim}{Y}|G_k) = \log L(\underset{\sim}{Y}|G_k;\ X_t|G_k,\ \underset{\sim}{\theta}_1,\ldots,\underset{\sim}{\theta}_T) - \sum_{k=1}^{k} \lambda_k(\sum_{t=1}^{T} X_t|G_k - 1)$

Taking the derivative of equation (39) with respect to $P(X_t|G_k)$ and setting to zero yields the following relationship.

(40) $\frac{\partial L}{P(X_t|G_k)} = \sum_{j=1}^{n_K} \frac{g_t(y_{jk};\underset{\sim}{\theta}_t)}{f(y_{jk})} - \lambda_k = 0$

Taking the derivative of equation (40) with respect to Θ_{it} and setting equal to zero yields the following relationship.

$$(41)\ \frac{\partial L}{\partial \Theta_{it}} = \sum_{k=1}^{K} \sum_{j=1}^{n_K} P(X_t|G_k)\ \frac{\partial g_t(y_{jk};\Theta_t)/\partial\Theta_{it}}{f(y_{jk})} = 0$$

Multiplying the relationship shown in (40) by $P(X_t|G_k)$ and summing over the t classes we see that the solution for the Lagrangian multiplier λ_k is n_k, the within group sample size.

Now, by Bayes theorem, the probability of observing a class X_t given y_{jk} is:

$$(42)\ P(X_t|y_{jk}) = \frac{P(X_t|G_K)g_t(y_{jk};\Theta_t)}{f(y_{jk})}$$

After multiplying equation (40) by $P(X_t|G_k)$ and rearrangement of terms we see that the MLE $\hat{P}(X_t|G_k)$ is of the form:

$$(43)\ \hat{P}(X_t|G_k) = \frac{1}{n_K} \sum_{j=1}^{n_K} P(X_t|y_{jk}).$$

In words, equation (43) says that the MLE of the mixing proportions $P(X_t|G_K)$ is given by the sample mean of the probability of observing class t given y_{jk}.

Substituting equation (42) into expression (41) yields the following.

$$(44)\ \sum_{k=1}^{K} \sum_{j=1}^{n_K} P(X_t|y_{jk})\ \frac{\partial L}{\partial \Theta_{it}} = 0;$$

that is, the MLE of the parameters Θ_{it} are the weighted averages of the MLEs $\partial L/\partial \underset{\sim}{\theta}_k$ which would arise from considering each component of the mixture separately. The weights are the probabilities of membership of Y_{jk} in each class t.

Estimation

We use the EM algorithm to estimate the parameters of the Latent Discriminant Model. The algorithm is simple and straightforward in that it closely follows the natural form of the maximum likelihood equations.

Let $P(\hat{X}_t|G_k)^{\ell}, \underset{\sim}{\theta}_t^{\ell}$ be the estimate at the ℓ^{th} iteration. The density of the vector y_{jk} in the t^{th} class may be expressed as

$$(46)\quad f(y_{jkt})^{\ell+1} = g(y_{jk};\underset{\sim}{\theta}_t)^{\ell} P(X_t|G_K)^{\ell}$$

The posterior probability of observing class t given the vector of observations y_{jk} is:

$$(47)\quad \hat{P}(X_t|y_{jk})^{\ell+1} = \frac{f(Y_{jkt})^{\ell+1}}{\sum_{t=1}^{T} f(y_{jkt})^{\ell+1}}$$

From (46) and (47) we may estimate the latent class proportions $\hat{P}(X_t|G_k)$ as

$$(48)\quad \hat{P}(X_t|G_k)^{\ell+1} = \frac{\sum_{j=1}^{n_k} \hat{P}(X_t|y_{jk})^{\ell+1}}{\sum_{j=1}^{n_k}\sum_{t=1}^{T} P(\hat{X}_t|y_{jk})^{\ell+1}} ;$$

In other words, the latent class proportion $\hat{P}(X_T|G_k)^{\ell+1}$ is the proportion of individuals that belong to latent class t from group k divided by the total proportion of individuals in group k. Equation (48) may also be written as

$$(49)\quad \hat{P}(X_t|G_k) = \frac{1}{n_k} \sum_{j=1}^{n_k} \hat{P}(X_t|y_{jk})$$

Although (49) is simpler to understand, (48) is used in computations to minimize roundoff error.

We now turn to the within class conditional maximum likelihood estimates for the manifest variables, denoted by the vector $\underset{\sim}{\Theta}_t$. The variables in $\underset{\sim}{\Theta}_t$ may come from any distribution of the exponential family. We will lay out the estimation process for variables from the normal and multinomial distributions. For the case of the normal distribution, the mean, $\Theta_{\bar{y}_{it}}$, for the i^{th} variable in the t^{th} latent class is:

$$(50)\quad \hat{\Theta}^{\ell+1}_{\bar{y}_{it}} = \frac{\sum_{k=1}^{K} \sum_{j=1}^{n_k} \hat{P}(X_t|y_{jk})(y_{ijk})}{\sum_{k=1}^{K} \sum_{j=1}^{n} \hat{P}(X_t|y_{jk})}$$

which is simply the mean, weighted by the probability of observing class t given the vector y.

The variance $\Theta_{var_{it}}$ for the i^{th} variable in the t^{th} latent class is

$$(51)\ \hat{\Theta}^{\ell+1} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{n_k} \hat{P}(X_t|y_{jk})(y_i-\hat{\Theta}_{\bar{y}_{it}})^2}{\sum_{k=1}^{K}\sum_{j=1}^{n_k} \hat{P}(X_t|y_{jk})}$$

For the multinomial case, let y_{im} denote the i^{th} multinomial variable with M categories. The estimate $\hat{\Theta}_{y_{imt}}$ for the proportion of the i^{th} variable at the m^{th} level within the t^{th} latent class is

$$(52)\ \hat{\Theta}^{\ell+1}_{y_{imt}} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{n_k} \hat{P}(X_t|y_{jk})(y_{imjk})}{\sum_{k=1}^{K}\sum_{j=1}^{n_k} \hat{P}(X_t|y_{jk})}$$

Using the estimates from (48) and (50), (51), or (52) as necessary, the process is repeated until convergence is achieved. Convergence is defined as

$$\{[P(\hat{X}_t|G_k)^{\ell+1},\ \hat{\Theta}^{\ell+2}]-[P(\hat{X}_t|G_k)^{\ell},\ \underset{\sim}{\rho}^{\ell}]\}<\underset{\sim}{\rho},$$

where $\underset{\sim}{\rho}$ is a constant vector of some arbitrarily small quantity.

The Basic Set

Due to the restrictions mentioned in expressions (35) and (36), the number of parameters we actually need to estimate is less than the number of parameters in the model. To illustrate this, assume we have a model with T latent classes, K groups, two polytomous variables with M categories, and two continuous variables distributed $N(\mu,\sigma^2)$. Due to the restrictions in expression (35), we need to estimate T-1 of the

$P(X_t|G_k)$; due to the restrictions in expression (36) we need only consider M-1 categories of each polytomous variable within each latent class. Thus our basic set will consist of $K(T-1)+2(M-1)T+(T)\cdot(2)$ parameters.

Goodness-of-fit

The statistic used in testing the goodness-of-fit of some hypothesized model m would be $-2\log(L_m/L_o)$. This statistic is distributed as a χ^2 random variable where the degrees of freedom is equal to the difference between the size of the basic set of L_a and the basic set of L_o. The likelihood ratio test may have extremely low power, therefore it should be used with caution.

Identification

The model will not be identified when the basic set is greater than N, where N is the number of unique profiles. If the basic set is less than or equal to N, we check identiability using a test suggested by McHugh (1956). If the Jacobian matrix of the parameters in the basic set is of full rank, then the parameters span the full latent space and the model is locally identified.

Some Additional Results

In addition to the maximum likelihood estimates, some results that arise from the latent discriminant model are given below.

The solution for the latent discriminant model was expressed in terms of the probability of observing a latent class given a group k, $P(X_t|G_k)$, and the conditional probabilities within each class. It would be informative to know the size of the classes, that is, the $P(X_t)$'s. From basic rules of probabilities we can write

$$(53)\quad P(X_t) = \sum_{k=1}^{K} P(X_t|G_k)P(G_k)$$

The recruitment probability for each class is

$$(54)\quad P(X_t|\underset{\sim}{y}) = \frac{P(\underset{\sim}{y}|\underset{\sim}{\theta}_t)\cdot P(X_t)}{\sum_{t=1}^{T} P(\underset{\sim}{y}|\underset{\sim}{\theta}_t)\cdot P(X_t)}$$

The probability of observing a group given a latent class is

$$(55)\quad P(G_k|X_t) = \frac{P(X_t|G_k)P(G_k)}{\sum_{k=1}^{K} P(X_t|G_k)P(G_k)}\ ,$$

which is simply the number of profiles in group k and latent class t divided by the number of profiles in latent class t.

Finally, the posterior probability of observing G_k given $\underset{\sim}{y}$ can be written as

$$(56)\quad P(G_k|\underset{\sim}{Y}) = \sum_{t=1}^{T} P(G_k|X_t)\cdot P(X_t|\underset{\sim}{Y})$$

Restricted Latent Discriminant Analysis

The majority of interesting applications involve the testing, in a confirmatory sense, of hypothesized structures. Various restrictions are permissible on the model which allows one to initiate a confirmatory test. Three basic types of restrictions may be placed on the model which are similar to the LISREL model (Jöreskog and Sörbom 1981), and the MLLSA model (Clogg 1979).

1. Fixed parameters that have been given assigned values.
2. Constrained parameters (equality restrictions) that are unknown but equal to one or more other parameters.
3. Free parameters that are not constrained and may take on any permissible value.

The degrees of freedom for a restricted model are equal to the number of nonredundant restrictions placed on the model plus the degrees of freedom of the unrestricted model (if identified). In evaluating the fit of the restricted t class model, we must compare it to the unrestricted t class model as virtually any set of fixed and free parameters will yield a better fit than the null model.

Synopsis

This chapter has laid out the latent discriminant model. For a model to have any applied value, the algorithm, as operationalized by the computer program should show a reasonable accuracy in recovering the true population parameters via the sample statistics. Chapter IV is devoted to a simulation study of estimating the parameters of the Latent Discriminant model via the EM algorithm as operationalized by the program LADI.

CHAPTER IV

SIMULATION

This chapter investigates the ability of the Latent Discriminant Model to capture a known structure. Specifically, a series of Monte Carlo simulation experiments are initiated in order to assess how well the Latent Discriminant Model will perform in practice. The chapter begins by giving the rationale for the simulation. The next section describes the purpose of the simulations. Next, the design for the simulation is specified, and each factor is explained in some detail. As the output from the simulation was voluminous, various summary variables were constructed and used in the analysis. These variables will be described in the course of the discussion. Finally, the results are given and limitations are discussed.

The Need for a Simulation

As stated at the end of Chapter II, for a model such as the Latent Discriminant Model to have any applied value, an efficient and reliable operationalization must exist. The algorithm must be efficient in the sense that it should move towards the local maximizer steadily and reliably, in the sense that it uncovers the true parameters under regular conditions. A simulation that tests the behavior of the program LADI across a number of parameter spaces provides information on both efficiency and reliability. By analyzing such performance criteria as the number of iterations and bias under alternative specifications, the efficiency and reliability of the Latent Discriminant Model can be better understood.

Purpose of the Simulation

The simulation primarily was designed to test whether the algorithm would recover a known specification, i.e., the true parameters. In addition, issues such as the behavior of the algorithm as a function of the number of variables, the number of classes, the number of groups, the separation of the class densities, and the sample size were of interest. The simulation was designed to shed light on all these issues.

The Design

The simulation consisted of a test of LADI over 108 different parameter spaces. These 108 cells were constructed by combinations of five factors: (1) the number of variables (three levels: 4, 5, and 6); (2) the number of classes (two levels: 2 and 4); (3) the number of groups (two levels: 2 and 4); (4) the discriminatory power of the variables, i.e., the separation of the classes as defined by the distance between the center of each class to every other class (three levels: "weak" discrimination, "medium" discrimination, and "large" discrimination); and (5) the sample size (three levels: "small" sample--100; "medium" sample--500; and "large" sample--100). Each of the factors is now explained in some detail.

The Number of Variables (NVAR)

This factor had three levels corresponding to (1) four variables, (2) five variables, and (3) six variables. Only three levels were chosen because (1) most latent class studies reviewed by the author utilized at most six variables, and (2) given time and resource constraints six variables produced a design that was reasonable and manageable from both a time and cost perspective. As the Latent Discriminant Model is designed to accommodate both continuous and categorical variables, a decision was made to treat the four variable case as a combination of two continuous and two categorical variables, the five variable case as a combination of two continuous and three categorical variables, and the six variable case as a combination of two continuous and four categorical variables.

The Number of Classes (NCLASS)

The number of classes was set at two and four. The reasons were similar to those stated previously for the number of variables factor. Two and four latent class models are common in the literature, and these levels allow one to add some variance to the design while keeping the size at a manageable level.

The Number of Groups (NGROUP)

The number of groups was set at two and four. In real applications, the grouping will usually represent some observed phenomena, typically of the binary type, e.g., brand purchased/not purchased, survival/death. Other types of phenomena may have more than two classifications, e.g., the perceptions of a subject with respect to an evoked set of brands. Most evoked sets range between 3 and 7 brands (Silk and Urban 1978). Again, due to time and space constraints, it was felt that considering four groups would be reasonable as well as representative of the latent class model appearing in the literature.

The Discriminatory Power of the Variables (DISC)

This factor had three levels: (1) "weak" discrimination between classes; (2) "medium" discrimination between classes; and (3) "strong" discrimination between classes. We will now describe the definition of discrimination for both the continuous and categorical variables. First, in the case of the continuous variables we assume normality. It is well known that the normal distribution is characterized by two parameters, the mean and the variance. Without loss of generality, in all of the simulations the variance of each continuous variable was set at 1.0.

Discrimination was defined in terms of the mean differences among the latent classes. Letting u_{ij} denote the mean for variable i within class j, define: (1) weak discrimination as $|u_{ij}-u_{ik}|=2$; (2) moderate discrimination as $|u_{ij}-u_{ik}|=4$; and (3) large discrimination was defined as $|u_{ij}-u_{ik}|=6$. Our definition of weak discrimination concurs with Hathaway (1983) who found that in his simulations, the mean separation for which mixture densities become unimodal is $|u_1-u_2|\leq 2$. Figures 4.1, 4.2, and 4.3 show two weakly separated, moderately separated and largely separated mixtures. Mean profiles of the continuous variables for the two class model for all three levels of discrimination are given in

Figure 4.4. Mean profiles for the continuous variable for the four class model for all three levels of discrimination are given in Figure 4.4.

In the simulations each categorical variable was assumed to have two levels. Letting p_{ijk} denote the probability of observing level i of categorical variable j in class k, define: (1) weak discrimination as $|p_{1jk}-p_{1j\ell}|=2$; (2) medium discrimination as $|p_{1jk}-p_{1j\ell}|=5$; and (3) large discrimination as $|p_{1jk}-p_{1j\ell}|=8$. Profiles for the categorical variables (for the two variable case) for the two class model are given in Figure 4.5. Profiles for the categorical variables for the four class model are given in Figure 4.5.

For the mixture proportions the probability of observing any latent class, given a group was the same across all classes.

The Sample Size (NCASE)

Maximum Likelihood estimation is based on large sample sizes. In applied research settings, what is "large" is very often a function of the available budget. It is important to know at what size the algorithm converges towards the single consistent local estimator. For this reason, sample sizes of 100, 500, and 1000 were chosen.

Monte Carlo Experiments

One hundred replications were generated within each cell. The following describes the procedure for generating the samples:

(1) For a given cell, 100 samples of the appropriate size were randomly generated. (The local independence phenomenon greatly eases the

task of sample size generation.) IMSL routines GGDA and GGNML were used to generate the sample. GGDA generates a discrete sample given population cell probabilities and GGNML generates normal deviates Z_i (i.e., $Z_i \sim N(0,1)$. Random normal deviates $N(M,S^2)$ may be obtained by transforming GGNML output according to $y_i=(z_i)S+M$.

(2) For each of the 100 samples, parameters were estimated using the algorithm described in Chapter III. Let $\hat{\underset{\sim}{\Theta}}{}_i^*$ stand for the parameter vector generated from the i^{th} replication. For each cell, values of $\hat{\underset{\sim}{\Theta}}{}_i^*$ were saved for i=1,100. Additionally it was noted whether the algorithm converged within 1000 iterations, and the number of iterations was recorded for each cell.

(3) After step (2) the following summary measures were computed. Let $\hat{\underset{\sim}{\Theta}}{}^{*\cdot}$ indicate the average: $\hat{\underset{\sim}{\Theta}}{}^{*\cdot} = \sum_{i=1}^{100} \hat{\Theta}_i^*/100$, where $\hat{\underset{\sim}{\Theta}}{}^{*\cdot}$ is the bootstrap estimate of the parameters. The standard errors of $\hat{\underset{\sim}{\Theta}}{}_i^{*\cdot}$ are calculated as (Efron, 1981):

$$(1)\ \hat{SE} = \left[\frac{1}{(100-1)} \sum_{i=1}^{100} (\hat{\Theta}_i^{*\cdot} - \hat{\Theta}^{*\cdot})^2\right]^{1/2}$$

These steps, 1-3, were carried out for each cell in the design. To evaluate the design, certain "performance" variables were also computed. These variables are described immediately below.

Performance Measures

The following performance measures were computed:

(1) The number of iterations for each replication. As the start values for the algorithm were the true parameter values, the number of

iterations within each cell would be a good indicator of the average rate of convergence. Hereafter this performance measure is denoted by ITER.

(2) The average bias for the means of the continuous variables. For each replication, the average bias was calculated. Hereafter this performance measure is denoted by CONBIAS.

(3) The average bias for the variance of the continuous variables statement. For each replication, the average bias in variance was calculated. Hereafter this performance measure is denoted by VARBIAS.

(4) The average bias for the categorical variables. For each replication, the average bias for nonredundant levels of the categorical variables was calculated. Hereafter this performance measure is denoted by CATBIAS.

(5) The average bias for the mixture proportions. For each replication, the average bias for nonredundant mixture proportions was calculated. Hereafter this performance measure is denoted by GBIAS.

(6) The average bias for the solution. This was defined as:

$$\sum_{\text{parameters}} (\text{BIAS}/\hat{\Theta}_i^*)/\text{\# of parameters.}$$

Hereafter this performance measure is denoted by TOTBIAS.

(7) The number of times bias > $(1/4\ \hat{SE})$/# of parameters. Hereafter this performance measure is denoted by ZKBIAS.

An analysis of variance was performed on each of the performance measures, using the five design variables as the independent factors. The BMDP2V package (BMDP 1981) was used for analysis. Table 1 gives a complete description of the design. The rows in this table correspond to the 108 cells of the simulation experiments.

Results

The results section will discuss the analysis of each dependent variable in turn. Table 2 lists the independent variables in the analysis. The analysis of variance results for each variable are then presented. As most of the three-way interactions in the analyses were significant, the interpretation strategy will focus on plots of these interactions. The section ends by presenting bootstrap results for a prototypical cell in the design, along with best-case and worst-case results for the same cell, to emphasize the need for bootstrapping in such models.

Analysis of ITER

An analysis of variance was performed, using the number of iterations as the dependent variable and the five design variables as the independent variables. The ANOVA results are presented in Table 3. All ANOVA results were generated using the BMDP2V package (BMDP 1981). We now interpret each significant three-way interaction.

NVAR x NCLASS x NGROUP. Figure 4.6 displays this interaction. As would be expected, ITER dramatically increase with the number of classes. The number of groups has little effect on ITER. Interestingly, ITER decreases in the four latent class case when the number of variables increases from 4 to 5. Note also that in the four class case, ITER is larger for six variables than for five variables. This is probably due to the fact that while both the five and six variable case impose some structure on the 4 class model, the 6 variable case involves a relatively higher dimensionality and therefore takes relatively longer to find the local maxima.

NVAR x NCLASS x NCASE. Figure 4.7 displays this interaction. There are two interesting points to note. First, at the 2 class level, there is not much difference in ITER. There is a slight decrease as NVAR increases and as NCASE increases. At the 4 class level, the 6 variable case with a sample size of 100 takes less iterations to converge than the 4 variable case, while the situation is reversed for the case of sample size of 1000. The best explanation for this is probably that the data are very sparse with samples of size 100 and 6 variables and 4 classes, which may lead to quick (and erroneous) convergence. The 1000 sample size allows the algorithm to search for the "true" structure in the data.

The second interesting feature is that a sample size of 500 takes longer to converge at the 4 class level than a sample size of 100 or 1000. As Poulsen (1981) has noted, this is probably due to the fact that error dominates at sample sizes of 100, while structure dominates at sample sizes of 1000, while at sample sizes of 500 neither error nor structure dominates, thus a longer search is required.

NVAR x NCLASS x DISC. Figure 4.8 gives the plot of the interactions. As DISC increases, the decrease in ITER is dramatic. The interpretation of the plot is obvious. If the classes are well separated, a quick and stable solution will be found. On the other hand, if they are not well separated, then an increase in NVAR helps.

NVAR x NGROUP x NCASE. Figure 4.9 plots the interactions. NGROUP seems to have little or no effect. Again, the interesting feature is the increased time it takes the sample of size 500 to reach convergence.

NVAR x NGROUP x DISC. Figure 4.10 plots the interactions. Again, NGROUP seems to have little or no effect. Increasing NVAR helps reduce ITER, except in the weak discrimination case. In the case of weak discrimination, the apparent unimodality of the mixtures probably increases the number of iterations required to separate the mixtures in the case of 6 variables. This is not to suggest that 5 variables are better; 6 variables probably allow the algorithm to find the true structure, albeit more slowly.

NCLASS x NGROUP x NCASE. Figure 4.11 plots the iteractions. The large sample size (NCASE=1000) shows no practical difference across class/group levels. Again, the interaction is significant due to the increase in ITER at the 4 class level for a sample size of 500.

NVAR x NCASE x DISC. Figure 4.12 plots the interactions for this case. At the "large" discrimination level, NVAR and NCASE seem not to matter, as the structure is very clear. At lower levels of discrimination, it is better to have a large sample size. NVAR again shows a quadratic effect, which is probably due to the increased dimensionality of the problem.

NCLASS x NCASE x DISC. Figure 4.13 plots the interactions. Again, the interesting phenomenon is the increase in ITER for the 4 class/weak discrimination case as NCASE increases. As NCASE increases, various local maxima cease to exist, and a single consistent estimator appears. It is intuitively appealing to posit that the lower number of iterations for the 4 class/weak discrimination show NCASE=100 is a result of local maxima. (This is borne out in the analysis of BIAS.)

NGROUP x NCASE x DISC. Though insignificant, this interaction adds no new information to our previous interpretation. For completeness, Figure 4.14 plots the interactions.

This completes our analysis of ITER. A discussion of this section will be postponed until after an analysis of the BIAS.

Analysis of CONBIAS and VARBIAS

Two analyses of variance were performed, using the average absolute bias in the means of the normlly distributed continuous variables (CONBIAS) and the average absolute bias in the variances of the continuous variables (VARBIAS) as dependent variables. The five design variables were the independent variables. Tables 4 and 5 present the ANOVA results for CONBIAS and VARBIAS, respectively. Again, we interpret each significant three-way interaction. The ANOVA results and plots of three-way interactions show since both CONBIAS and VARBIAS show the same pattern of BIAS, they will be discussed together. In the discussion to follow when the term BIAS is used it refers to both CONBIAS and VARBIAS.

NVAR x NCLASS x NCASE. The interaction plots for VARBIAS and CONBIAS are in Figures 4.15 and 4.20, respectively. It is clear from these figures that NVAR has little effect on either type of bias. However, in several, the bias is reduced by an increase in NCASE and a decrease in NCLASS.

NVAR x NCLASS x DISC. The interaction plots are shown in Figure 4.16 for VARBIAS and Figure 4.21 for CONBIAS. Again, NVAR has little or no effect. An interesting interaction occurs between the 2 and 4 class models, however. The 2 class model requires only medium discrimination to reach a significant reduction in BIAS, while the 4 class model requires large discrimination to achieve a significant reduction.

NCLASS x NGROUP x DISC. Figures 4.17 and 4.22 plot the interactions for VARBIAS and CONBIAS, respectively. The findings are the same as in the NVAR x NCLASS x DISC case. NGROUP has little or no effect, the 2 class model requires only discrimination for a significant reduction in BIAS, whereas the 4 class model requires large discrimination.

NVAR x NCASE x DISC. Figures 4.18 and 4.23 plot the interactions for VARBIAS and CONBIAS, respectively. Again, NVAR has little or no effect, except for the 6 variable/weak discrimination case. With large samples, the results are good regardless of the discrimination level. With small sample sizes, however, an increase in the level of discrimination helps dramatically.

NCLASS x NCASE x DISC. Figures 4.19 and 4.24 plot the interactions for VARBIAS and CONBIAS. The 4 class/weak discrimination case shows high BIAS regardless of the sample size. Otherwise, the large sample solutions are robust, and solutions improve with increased discrimination at smaller sample sizes.

Summary. NGROUP and NVAR have little or no effect on BIAS. Large discrimination is preferred. For low bias, an increase in NCASE proportional to an increase in NCLASS is necessary in order to reduce bias.

Analysis of CATBIAS

An analysis of variance was performed, using the average absolute bias of the categorical variables as the criterion measure and the five design variables as the independent variables. Table 6 presents the ANOVA results. We now interpret the significant three-way interactions.

NVAR x NCLASS x NCASE. Figure 4.25 plots the interactions. The number of variables has little effect on BIAS. At the 2 class level, BIAS is tolerable at sample sizes of 500 (BIAS≃.02), while at the 4 class level a BIAS of this same magnitude is reached at a sample size of 1000.

NVAR x NCLASS x DISC. Figure 4.26 plots the interactions. NVAR again has little effect on BIAS. For 2 classes, BIAS is tolerable at medium discrimination (BIAS≃.03), while at the 4 class level large discrimination helps.

NVAR x NCASE x DISC. Figure 4.27 plots the interactions. With an increase in NVAR, larger sample sizes and high discrimination are needed to achieve the same accuracy.

NCLASS x NCASE x DISC. Figure 4.28 plots the interactions. There is a distinct difference between the 2 and 4 class solutions. At the 2 class level, samples of size 500 achieve a tolerable amount of BIAS (≃.05), while at the 4 class level, samples of size 1000 with medium discrimination are necessary to achieve the same amount of BIAS.

Summary. The number of groups has no effect on CATBIAS. As NCLASS increases, a larger sample size is necessary. The greater the discrimination between classes, the less the BIAS.

Analysis of GBIAS

GBIAS is the average absolute bias in the mixing proportions. The analysis of variance results are presented in Table 7. As these proportions play a crucial part in interpreting a solution, we will judge all the interactions assuming that an acceptable level of BIAS is lower than .03. The significant three-way interactions are now interpreted.

NVAR x NCLASS x NCASE. Figure 4.29 plots the interactions. NVAR has little effect on BIAS. For the 2 class model, an acceptable level of BIAS is reached with samples of size 500, while the 4 class model requires samples of size 1000.

NVAR x NCLASS x DISC. Figure 4.30 plots the interactions. Again, NVAR has little effect on the solution. Regardless of NCLASS, an acceptable level of BIAS is achieved at medium discrimination. An intriguing result is the decrease in BIAS for the 4 class/large discrimination case over the 2 class/large discrimination case.

NVAR x NGROUP x NCASE. Figure 4.31 plots the results for this interaction. Again, NVAR has little effect. With 2 groups, a sample size of 500 is adequate for meeting our BIAS standard of .03 or lower. For the 4 group case, a sample size of 1000 is required to meet this standard.

NCLASS x NGROUP x DISC. Figure 4.32 plots the results for this interaction. The interaction is very pronounced. At the 4 group level, we do not meet the BIAS standard for any level of NCLASS or DISC. At both levels of NGROUP, BIAS is reduced as DISC and NCLASS are increased. This somewhat intriguing result is probably due to the scaling of the mixing proportions. As we increase the number of classes, the expected proportions get smaller, allowing for less variation. Simultaneous increase in the level of discrimination further reduces the variance.

NCLASS x NCASE x DISC. The interactions are plotted in Figure 4.33. At a sample size of 100, we do not meet the BIAS standard. At sample sizes of 500 and 1000, we meet the BIAS standard for large and medium discrimination. Again, the interesting result is that BIAS is reduced as NCLASS and DISC are increased.

Summary of GBIAS. Sample sizes of 500 seem necessary. As the number of classes and the level of discrimination increases BIAS decreases. This is reassuring as an increase in NCLASS leads to mixing proportions that are smaller in absolute value.

Analysis of ZKBIAS

ZKBIAS measures the proportion of parameters in a solution for which BIAS > 1/4 STANDARD ERROR. This is a test suggested by Efron (1981). If BIAS > 1/4 STANDARD ERROR, then the amount of BIAS in the parameter may be considered "large" for the distribution of the parameter, and methods of correction should be used. As the analysis will show, ZKBIAS may be considered large for the LADI solution by Efron's standard. This is due to the fact that we have extremely tight standard errors for our solutions. The analysis of variance results are presented in Table 8. Though some of the effects are significant, the mean squares for the effects are too low to warrant serious consideration. Two three-way interactions are significant: (1) NVAR x NCLASS x DISC and (2) NVAR x NCASE x DISC. These interactions are plotted in Figures 4.34 and 4.35. It is obvious that ZKBIAS cannot be reduced by a researcher by controlling factors. Bootstrapping leads to a considerable reduction in ZKBIAS, as will be demonstrated in the final section.

Analysis of TOTBIAS

TOTBIAS is defined as $\sum(\text{BIAS}/\hat{\Theta})$/# of parameters, thus it measures the overall BIAS in a solution. An analysis of variance was performed, with TOTBIAS as the dependent variable and the five design factors as the independent variable. The ANOVA results are reported in Table 9. Only significant three-way interactions are interpreted.

NVAR x NCLASS x NGROUP. TOTBIAS increases with an increase in all, NVAR, NCLASS, and NGROUP. The interactions are plotted in Figure 4.36.

NVAR x NCLASS x NCASE. The interactions are plotted in Figure 4.37. TOTBIAS increases with an increase in NVAR, NCLASS, and NCASE. TOTBIAS reduces dramatically with an increase in NCASE.

NVAR x NCLASS x DISC. Figure 4.38 plots the interactions. At the 2 class level TOTBIAS shows a slight increase with an increase in NVAR while at the 4 class level, TOTBIAS shows a decrease with an increase in NVAR. An increase in DISC has a greater effect in reducing BIAS at the 4 class level as compared to the 2 class level.

NVAR x NGROUP x NCASE. Figure 4.39 plots the interactions. Increasing NCASE to 1000 reduces TOTBIAS dramatically. Increasing NVAR increases TOTBIAS at NGROUP = 2 but does not have the same effect at NGROUP = 4. An increase in the number of groups increases TOTBIAS.

NCLASS x NGROUP x DISC. Figure 4.40 plots the interactions. An increase in NCLASS and NGROUP increases TOTBIAS. An interesting feature is the higher TOTBIAS at the high level of DISC over the medium level of DISC. This can be attributed to the scaling of the categorical variables. As the conditional probability is .1 for observing the second level of the categorical variable, the ratio of bias to the value of the parameter is high.

NVAR x NCASE x DISC. Figure 4.41 plots the interactions. TOTBIAS increases with an increase in NVAR. A sample size of 500 with medium discrimination seems necessary to reach an acceptable level of TOTBIAS.

NCLASS x NCASE x DISC. Figure 4.42 plots the interactions. TOTBIAS increases with an increase in NCLASS, but this effect disappears as an NCASE and DISC increase simultaneously.

NGROUP x NCASE x DISC. Figure 4.43 plots the interactions. BIAS increases with an increase in NGROUP, but this effect disappears with an increase in NCASE and DISC.

NCLASS x NGROUP x NCASE. Figure 4.44 plots the interactions. BIAS increases with an increase in NGROUP and NCLASS, though NCLASS increases BIAS more dramatically than NGROUP.

Summary of TOTBIAS. A sample size of 500 seems necessary for controlling BIAS. As the number of classes increases, the sample size must increase proportionately. NGROUP has no noticeable effect on BIAS.

Some Bootstrap Results

As the previous results show, various problems plague the LADI solution in practical applications. BIAS due to sampling, lack of convergence to the consistent local estimator (again due to sample fluctuations), and the influence of the level of discrimination between the "true" mixtures are the sources of main concern. This leads us to Bootstrapping. Bootstrapping is simply a resampling method (Efron 1981). Given the data at hand, random subsamples are drawn independently and parameters are estimated for each subsample. The Bootstrap estimates are then obtained by methods described earlier in this chapter. These estimates, as will be shown, largely reduce the BIAS. Cell no. 89 from the design in Table 1 is used to illustrate the advantage of Bootstrapping. Cell 89 consists of a 2 class/4 group/6 variable LADI model with a sample size of 500. Table 10 gives the "true" parameters, the Bootstrap estimates, the standard error, the minimum value attained by the parameter upon all simulation, the maximum value attained by the parameter over all simulation, the BIAS in the Bootstrap estimates, and the "worst-case" BIAS, i.e., the large deviation from the "true" value. A comparison of Bootstrap BIAS to worst case BIAS shows a reduction in BIAS in the neighborhood of 90%. Bootstrap BIASES are very respectable. These results suggest that LADI can be used adequately for estimation in conjunction with an adequate resampling plan.

Summary and Limitations

Limitations

(1) The design aims for 100 repetitions within each cell. Due to non-convergence and system limitations, certain cells had fewer repetitions. The number of repetitions within each cell is listed in Table 11.

(2) The mixing proportions were uniform. Well discriminating groups may reduce BIAS and speed convergence.

(3) The simulation did not test the robustness to misspecification of distributions, a situation that may arise in application.

Summary

The Monte Carlo experiments show that LADI performs satisfactorily under regular conditions. It seems that a major prerequisite for accurate results is "large" sample sizes. A sample size of 250 per latent class seems adequate for accurate results. The number of groups adds little or no BIAS, an encouraging signal for applications. With sample sizes smaller than 250, or weak discrimination, Bootstrapped estimates can greatly improve reliability.

This concludes the Simulation chapter. The chapter detailed the results of a Monte Carlo experiment designed to test the adequacy of the Latent Discriminant Model. Performance measures were analyzed and sources of variance noted. In Chapter V we discuss empirical applications of the LADI model.

FIGURE 4 1 -- WEAK DISCRIMINATION

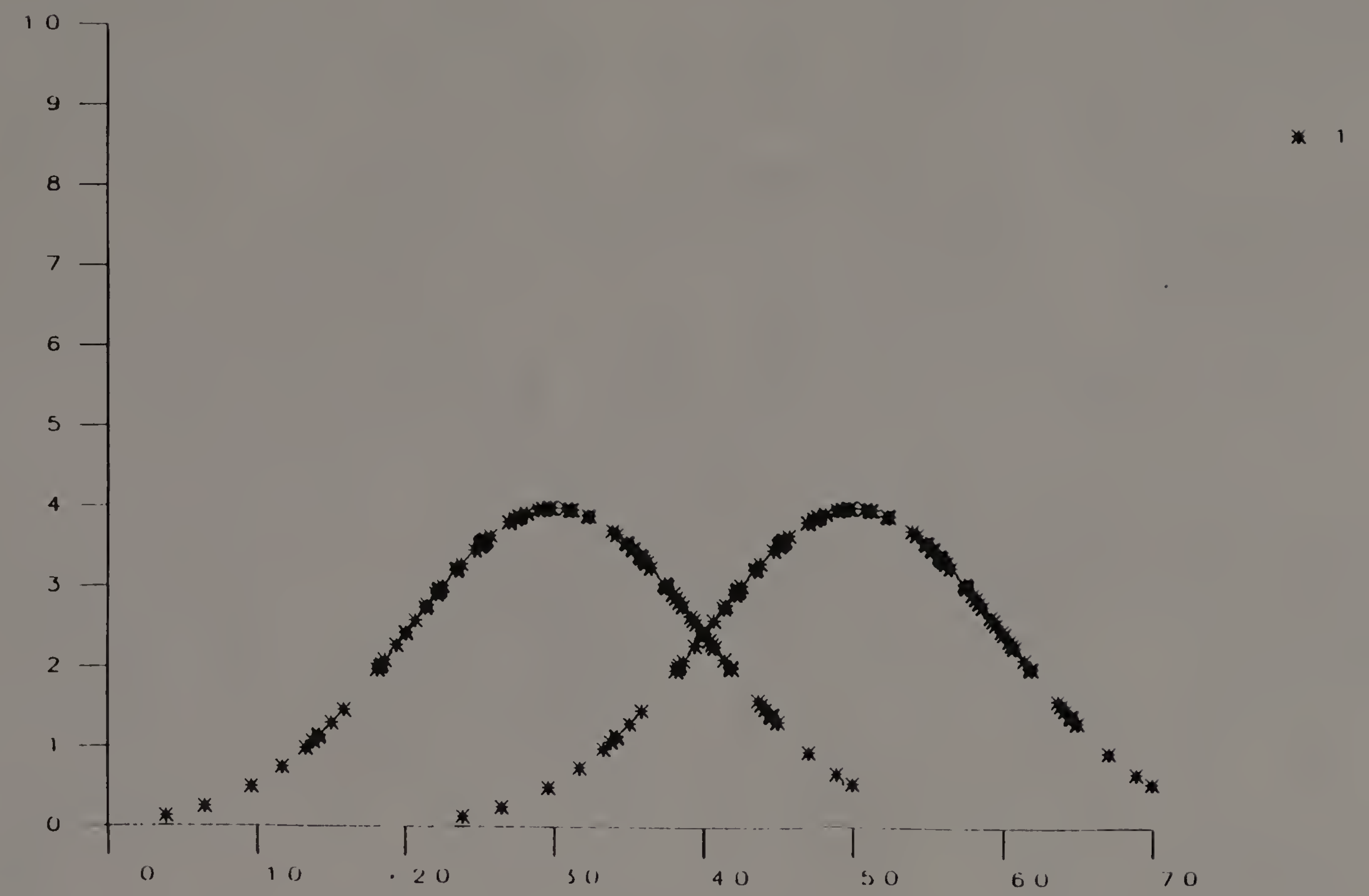

FIGURE 4.2 - MEDIUM DISCRIMINATION

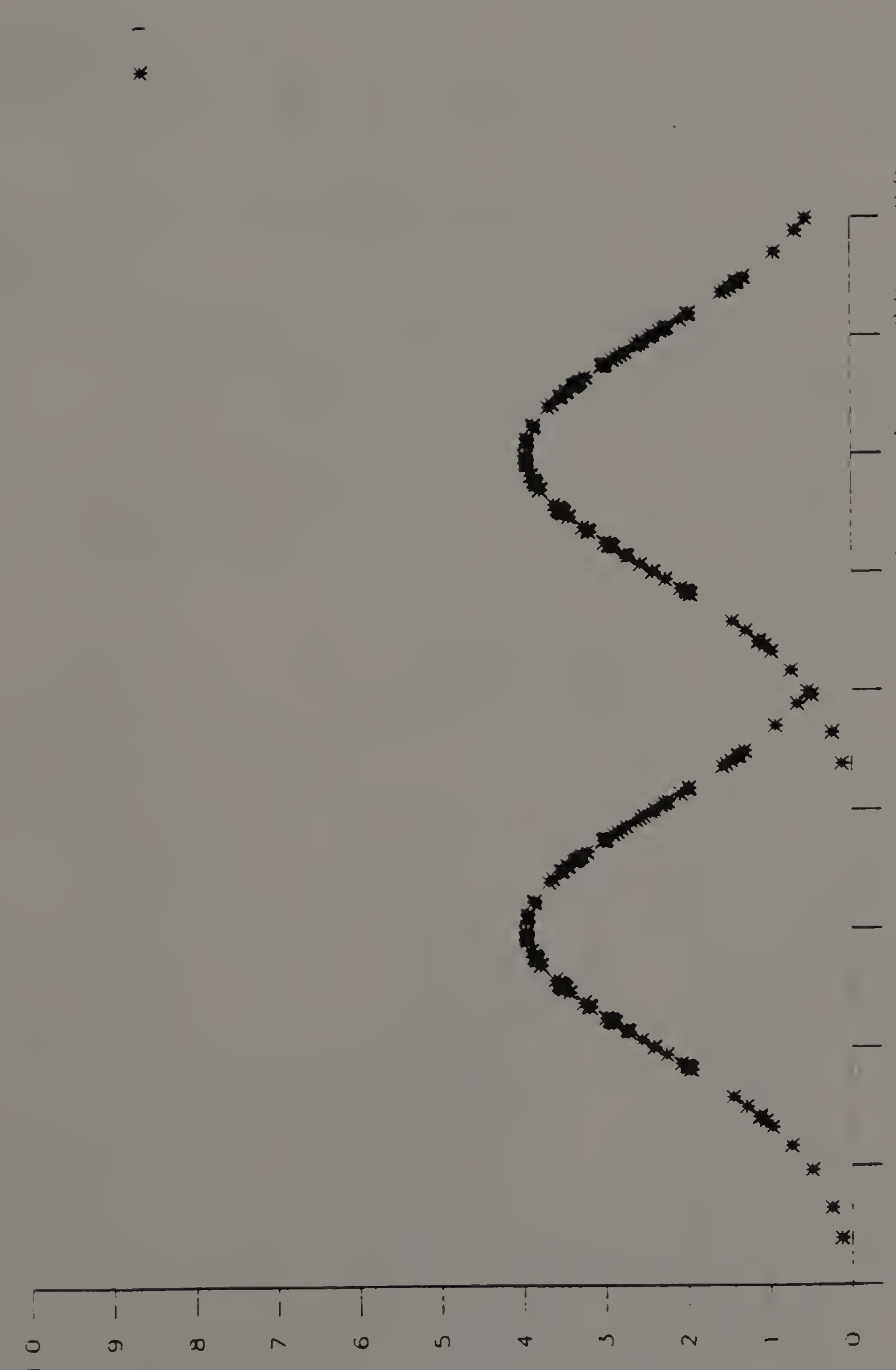

FIGURE 4.3 — LARGE DISCRIMINATION

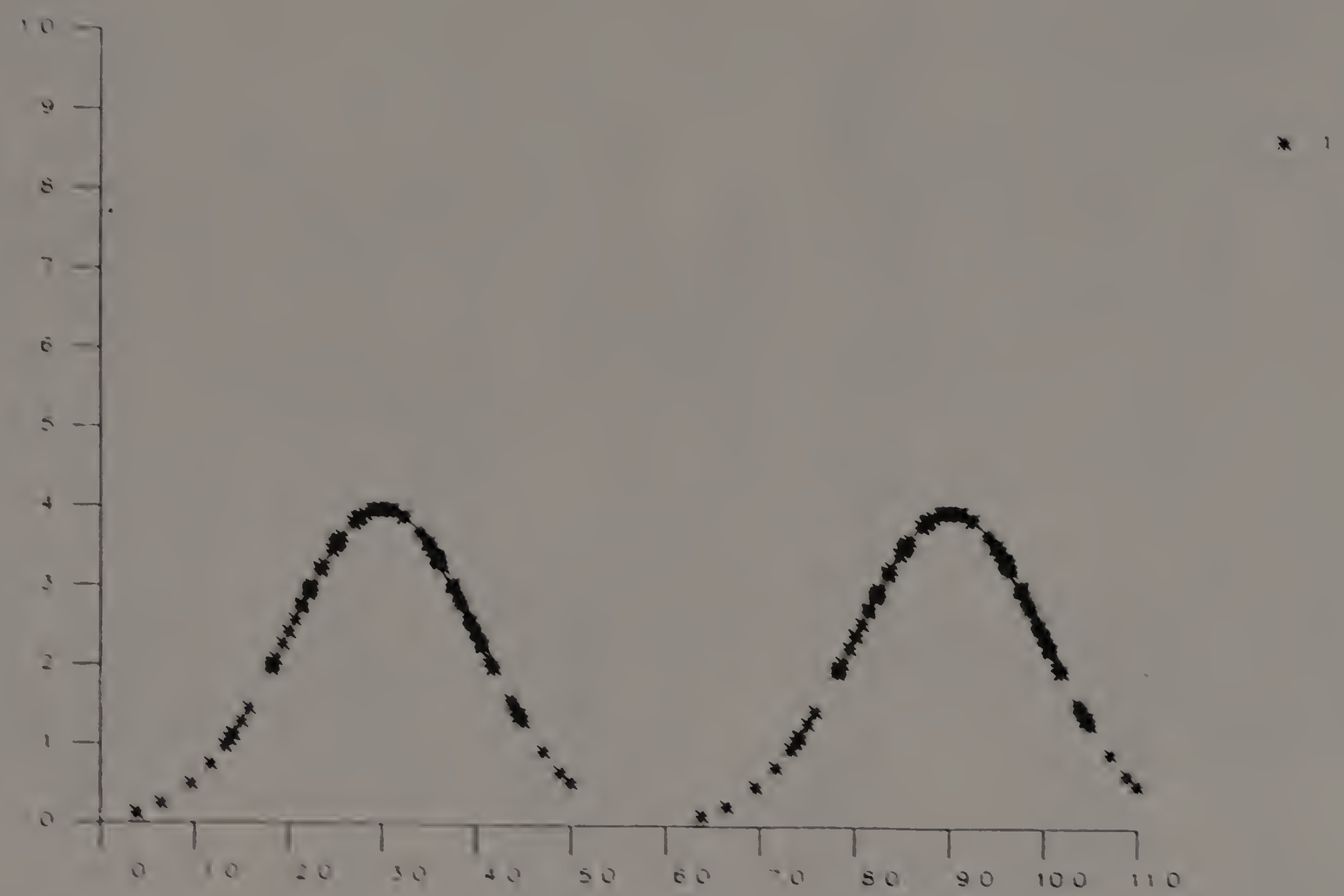

Figure 4.4

WEAK DISCRIMINATION, 2 CLASS CONTINUOUS VARIABLES

Figure 4.4 (continued)

MEDIUM DISCRIMINATION, 2 CLASS/CONTINUOUS VARIABLES

Figure 4.4 (continued)

LARGE DISCRIMINATION, 2 CLASS/CONTINUOUS VARIABLES

Figure 4.4 (continued)

WEAK DISCRIMINATION/4 CLASS/CONTINUOUS VARIABLES

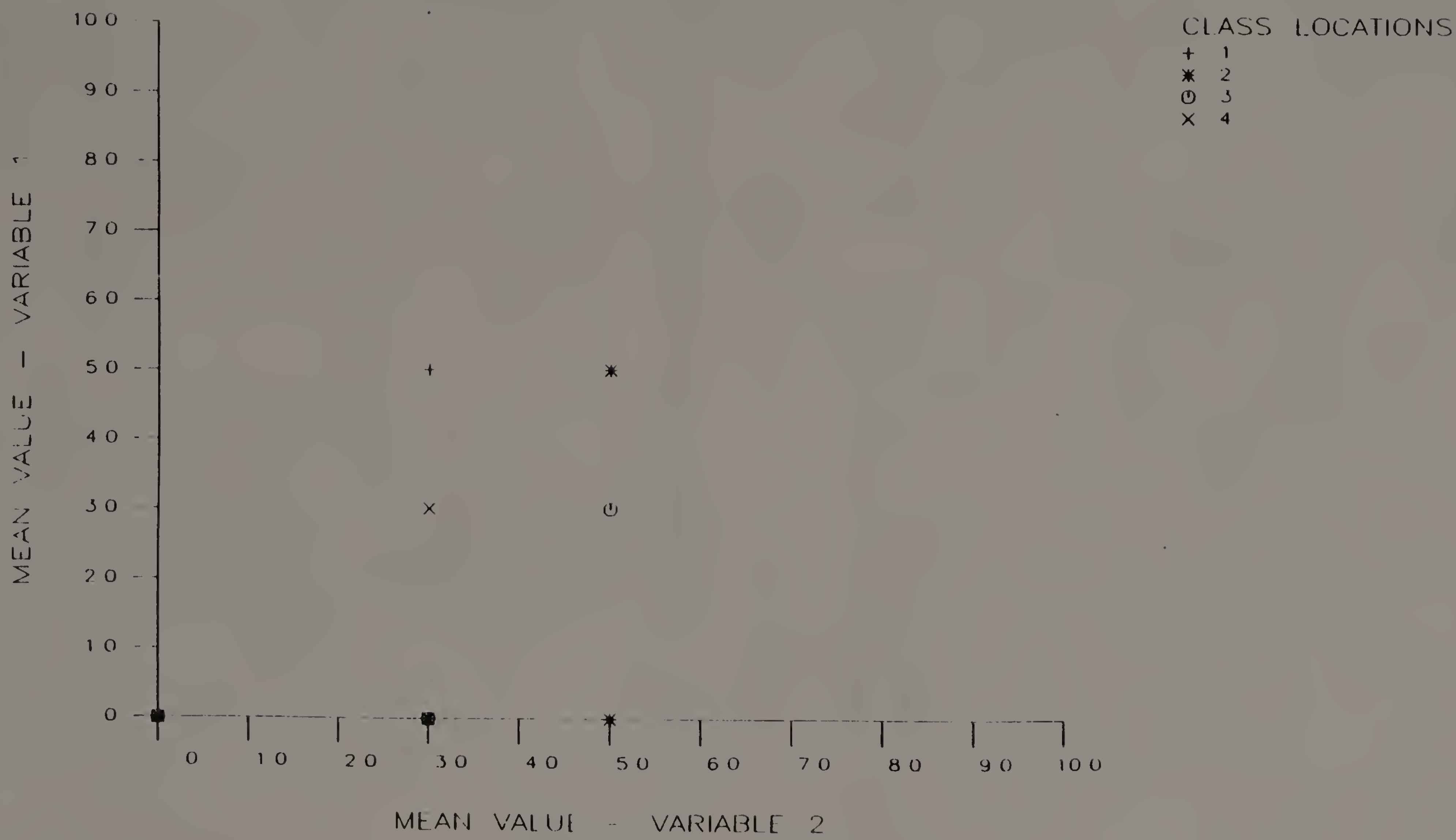

Figure 4.4 (continued)

MEDIUM DISCRIMINATION, 4 CLASS/CONTINUOUS VARIABLES

Figure 4.4 (continued)

LARGE DISCRIMINATION/4 CLASS/CONTINUOUS VARIABLES

Figure 4.5

WEAK DISCRIMINATION/2 CLASS/NOMINAL VARIABLES

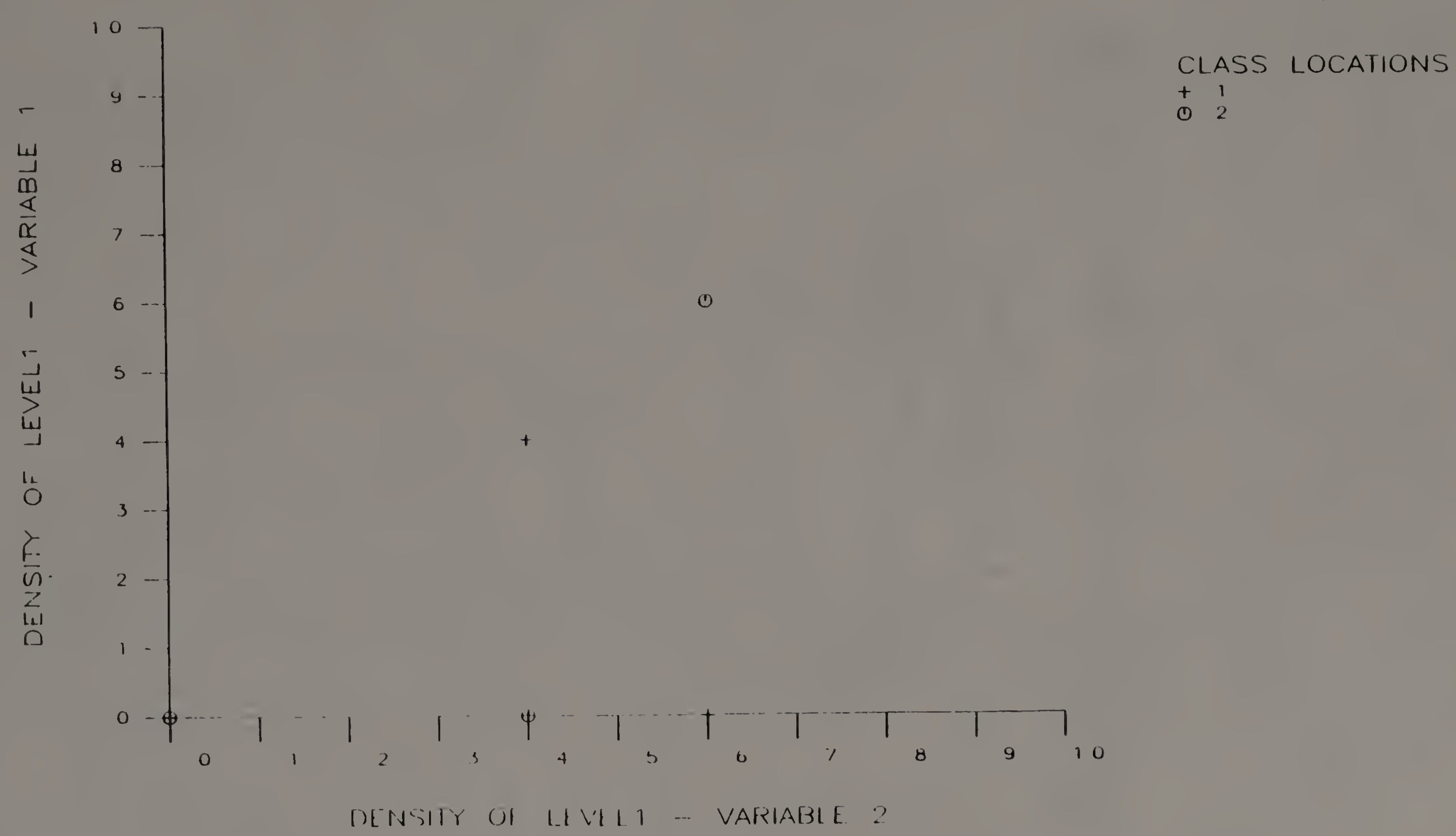

Figure 4.5 (continued)

MEDIUM DISCRIMINATION/2 CLASS/NOMINAL VARIABLES

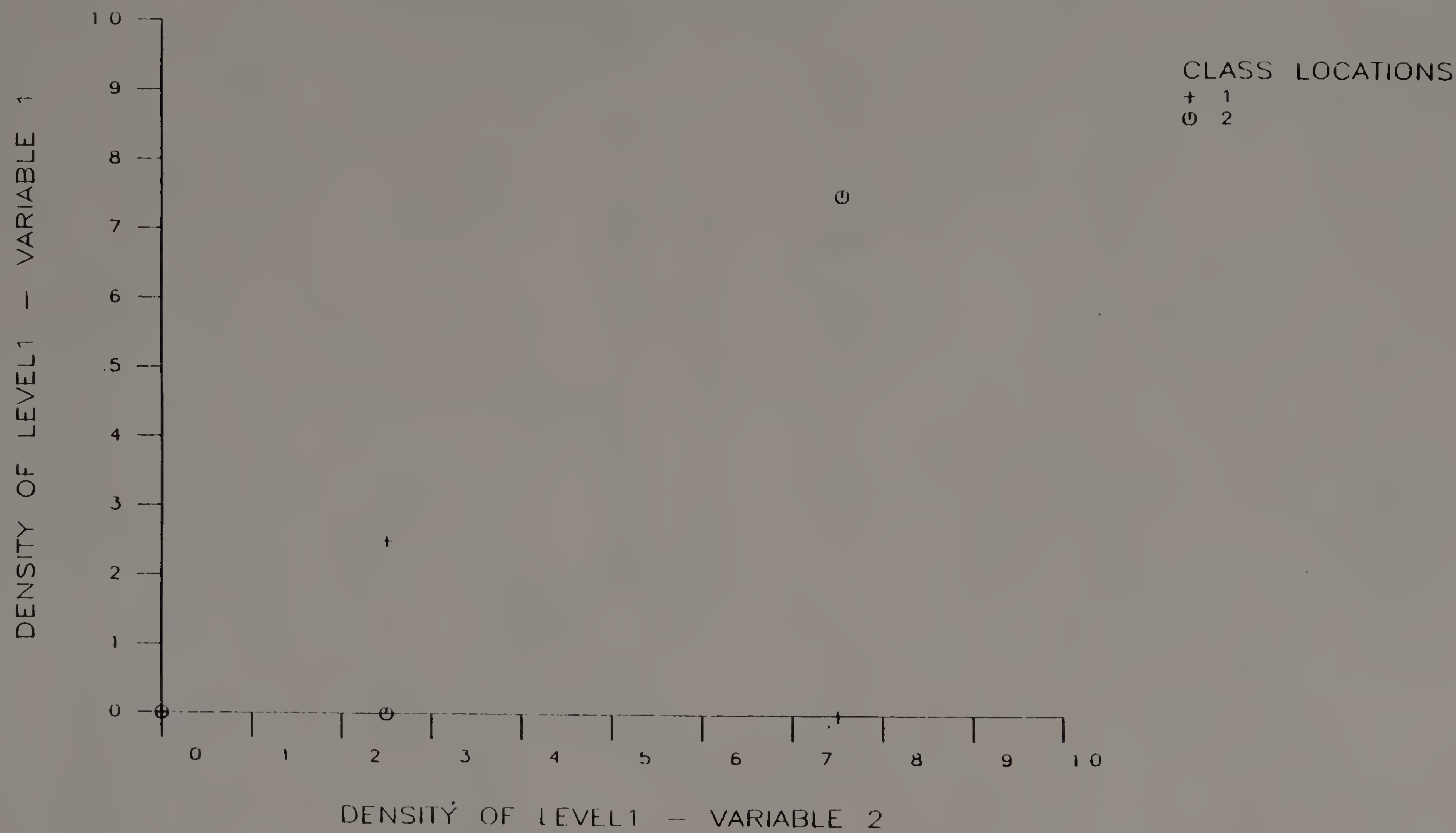

Table 4.5 (continued)

LARGE DISCRIMINATION/2 CLASS/NOMINAL VARIABLES

Figure 4.5 (continued)

WEAK DISCRIMINATION/4 CLASS/NOMINAL VARIABLES

Figure 4.5 (continued)

MEDIUM DISCRIMINATION/4 CLASS/NOMINAL VARIABLES

Figure 4.5 (continued)

LARGE DISCRIMINATION/4 CLASS/NOMINAL VARIABLES

FIGURE 4.6 TER BY NVAR*NCLASS*NGROUP

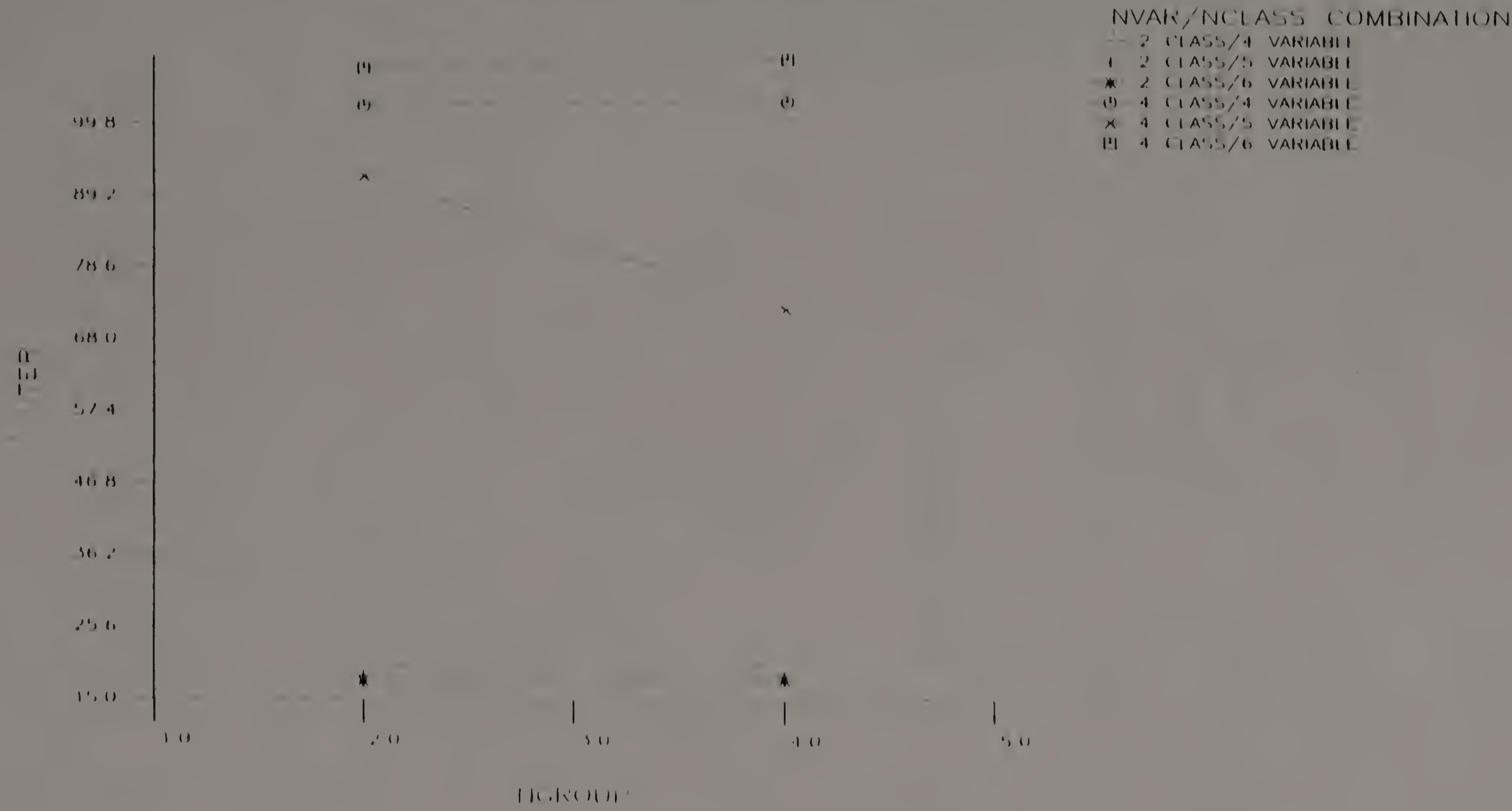

FIGURE 4.7 ITER BY NVAR*NCLASS*NCASE

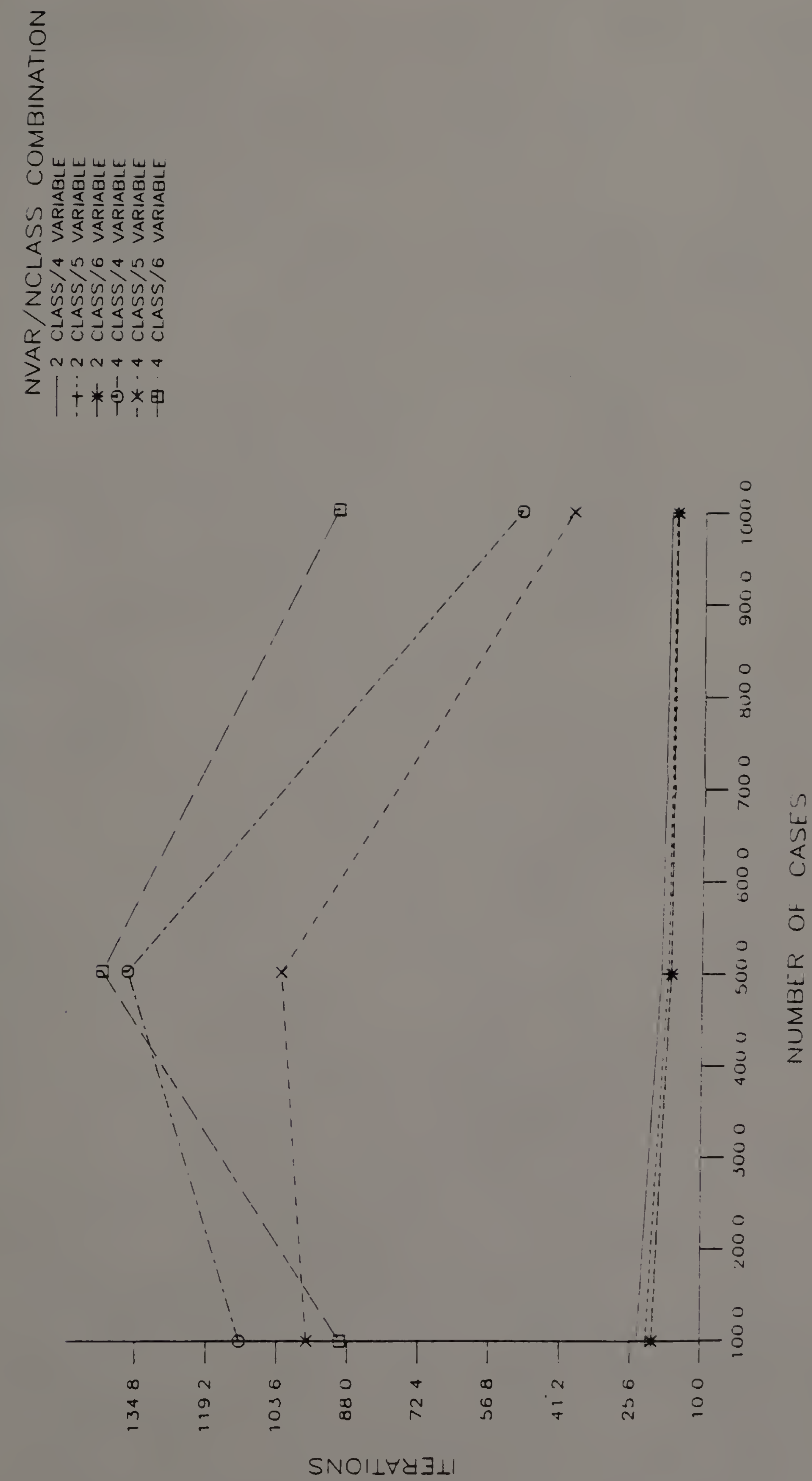

FIGURE 4.8 ITER BY NVAR*NCLASS*DISC

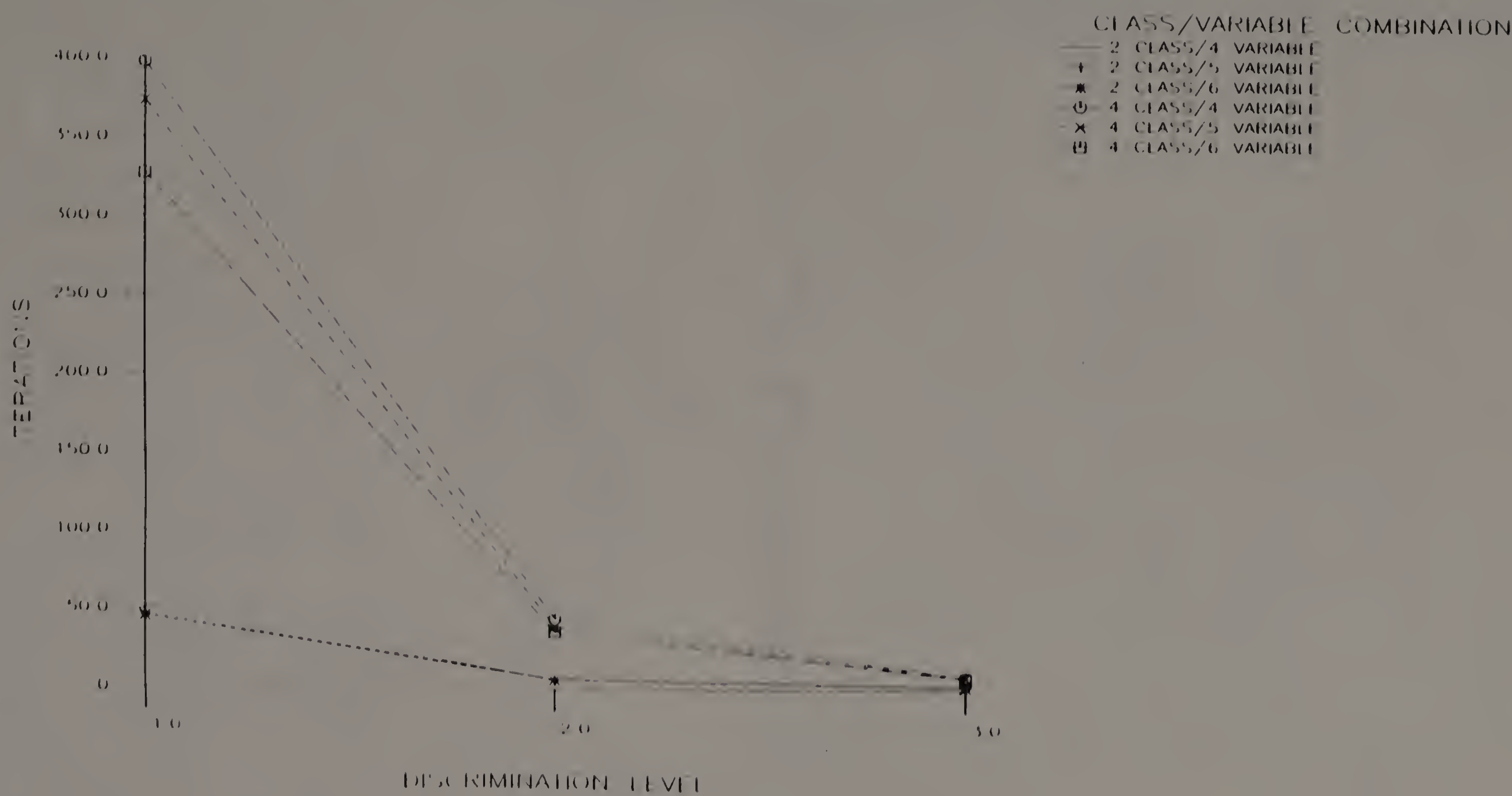

FIGURE 4.9 ITER BY NVAR*NGROUP*NCASE

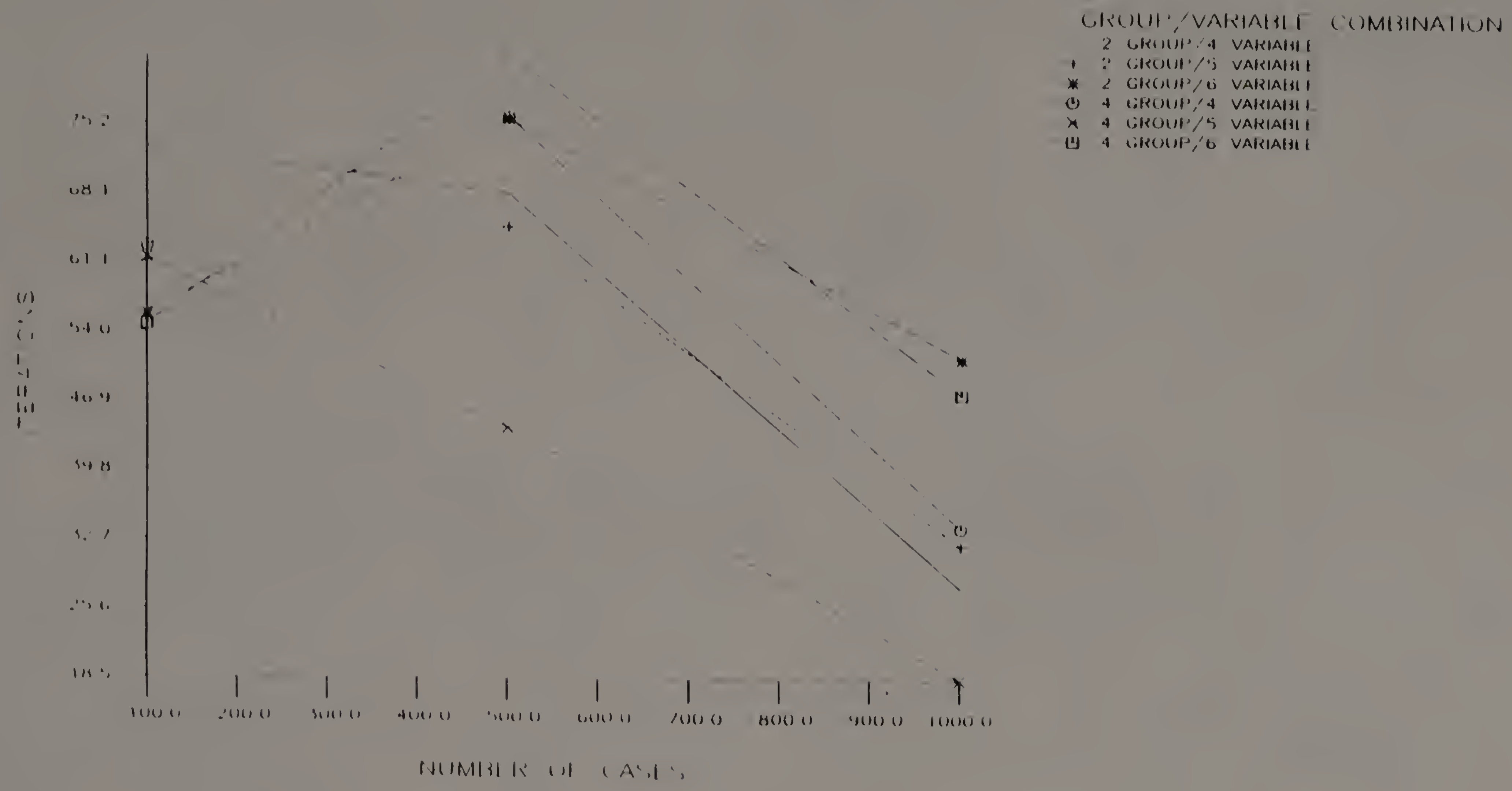

FIGURE 4.10 ITER BY NVAR*NGROUP*DISC

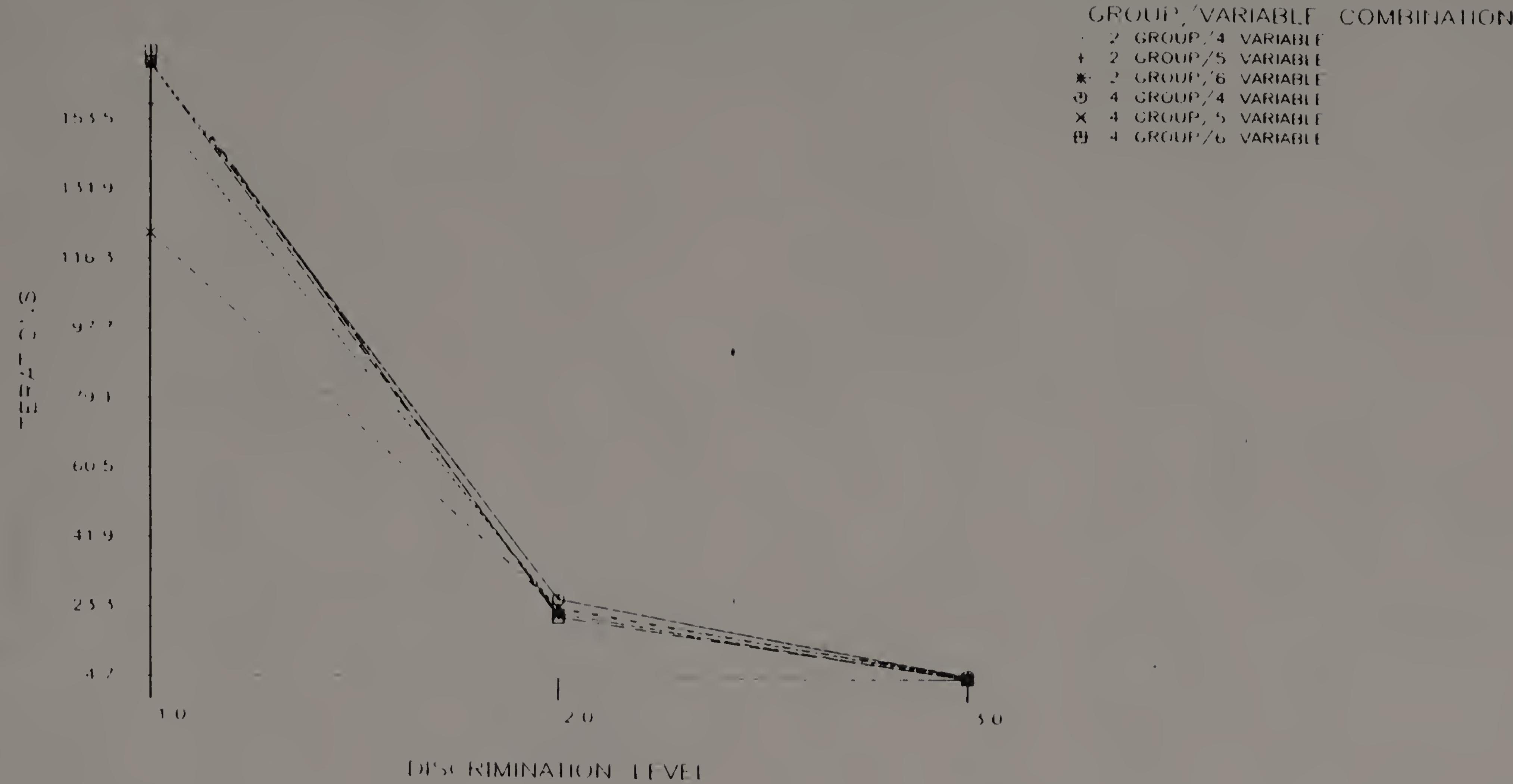

FIGURE 4.11 ITER BY NCLASS*NGROUP*NCASE

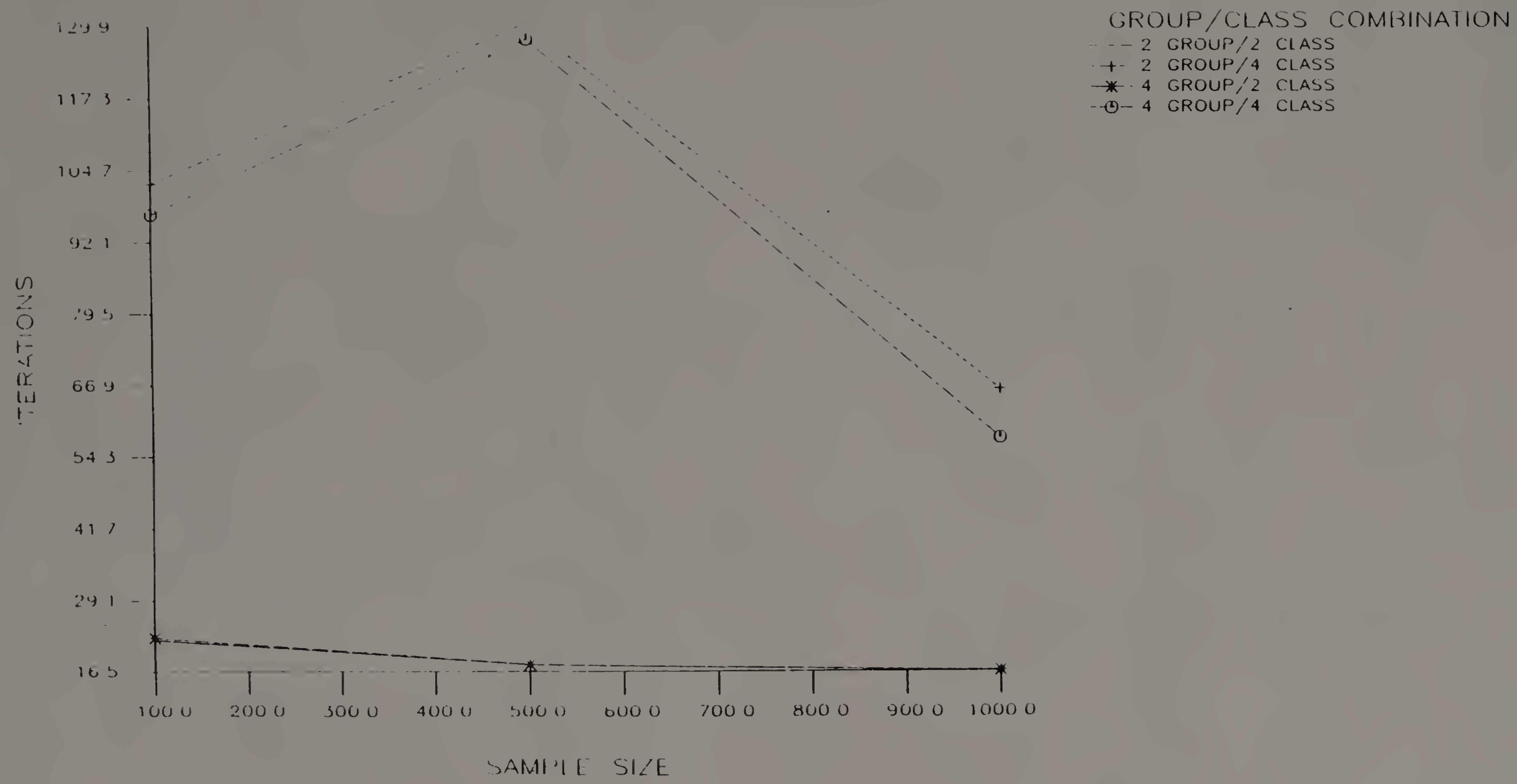

FIGURE 4.12 ITER BY NVAR*NCASE*DISC

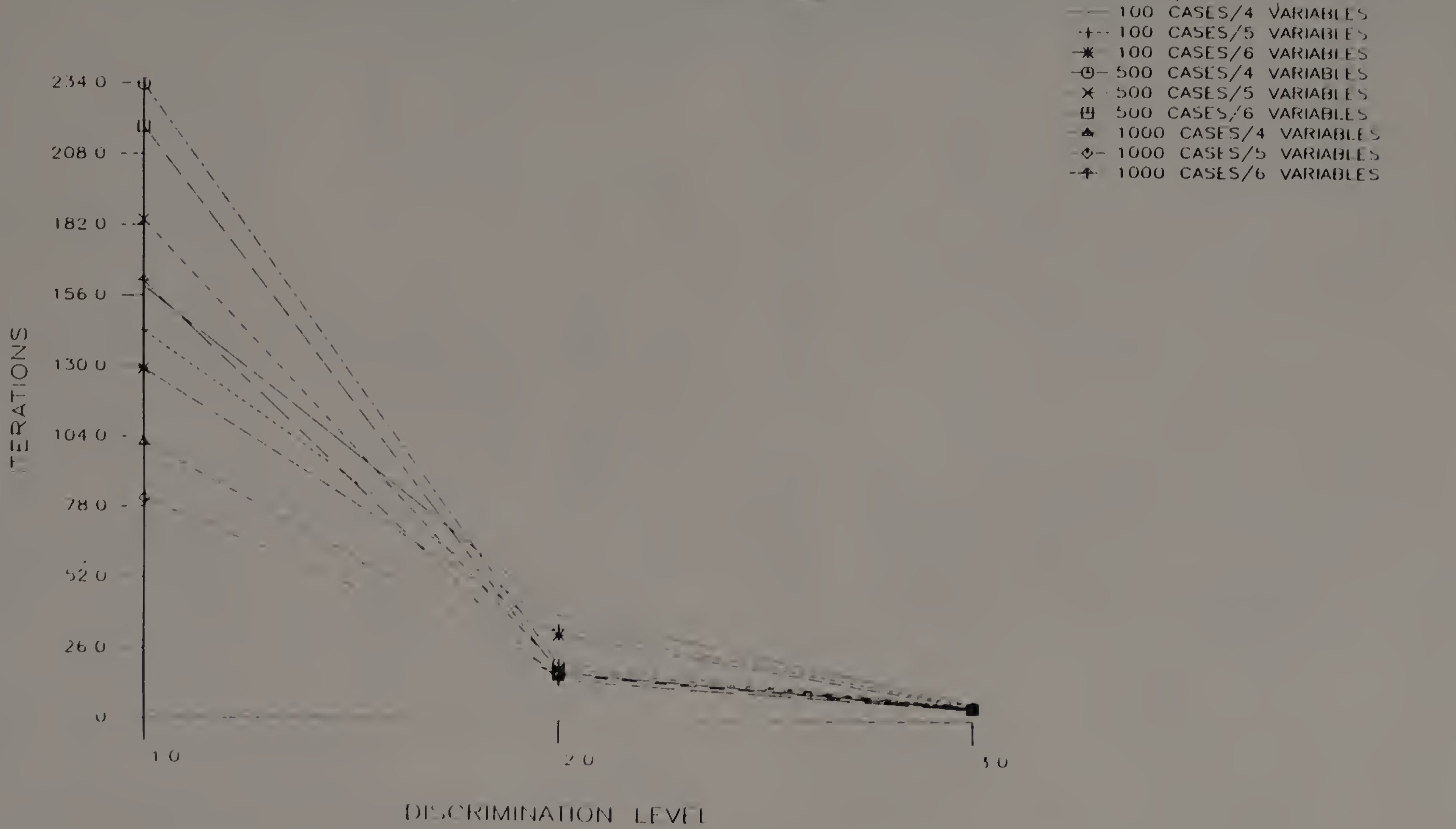

FIGURE 4.13 ITER BY NCLASS*NCASE*DISC

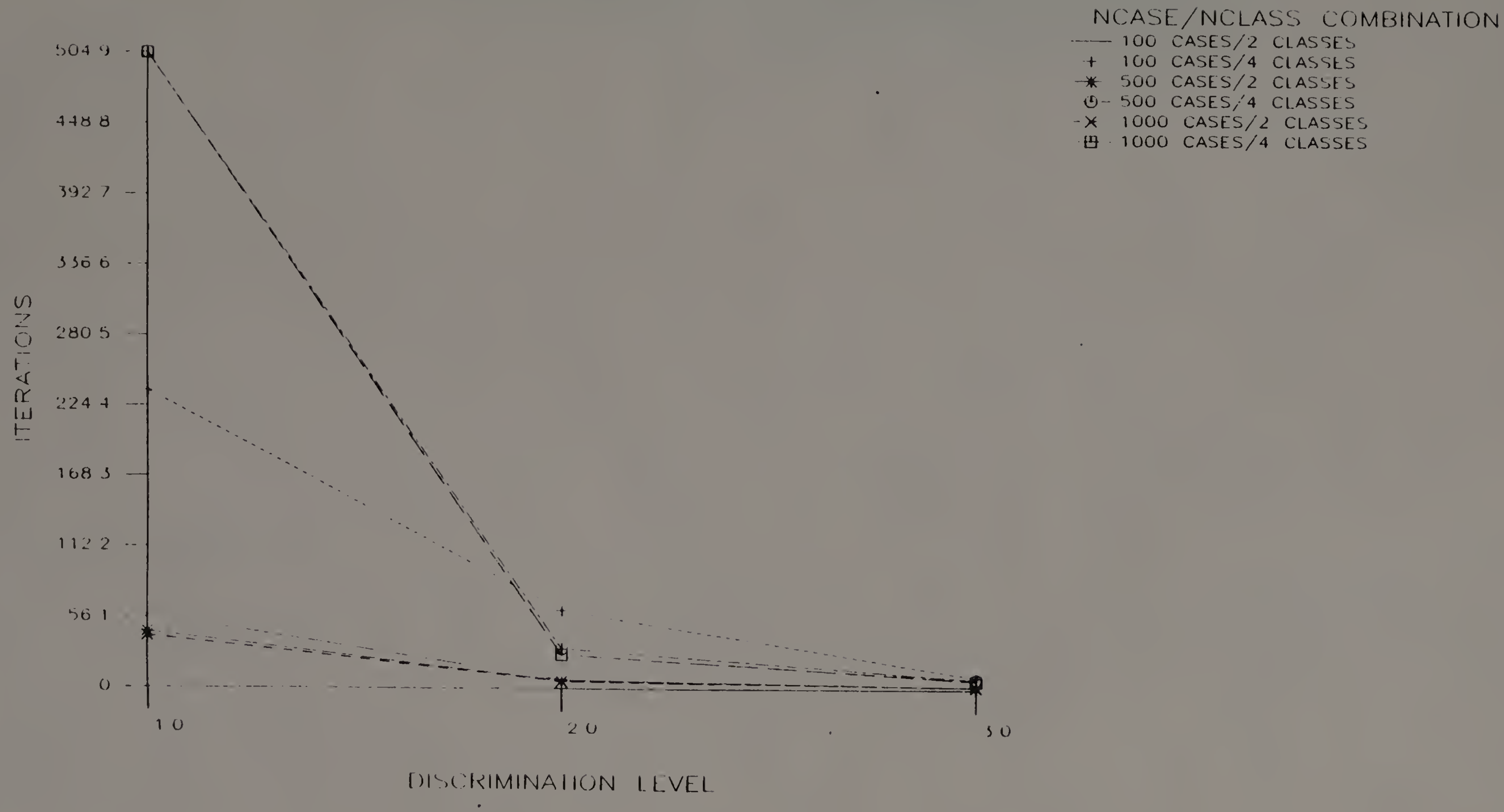

FIGURE 4.14 ITER BY NGROUP*NCASE*DISC

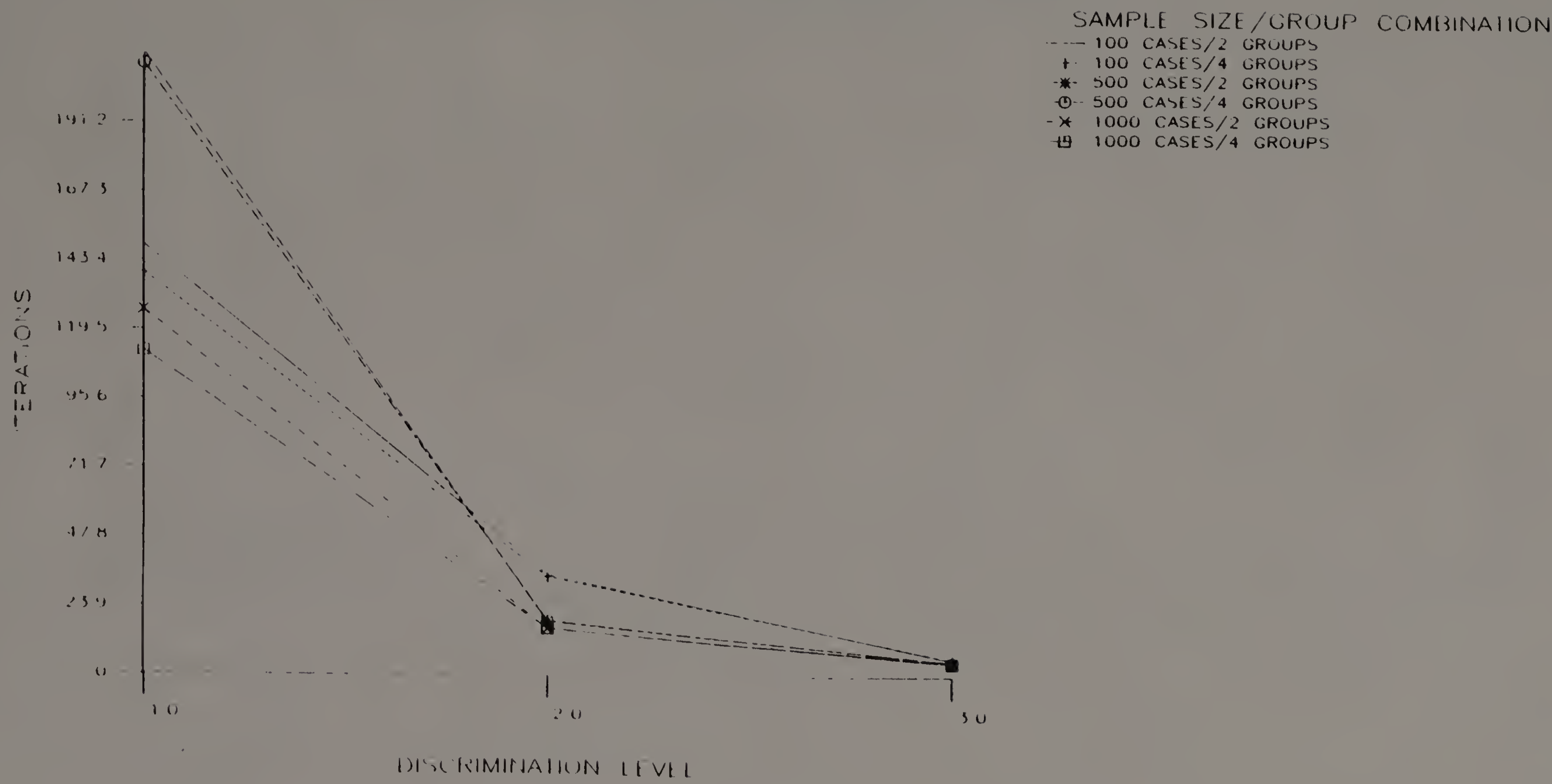

FIGURE 4.15 VARBIAS BY NVAR*NCLASS*NCASE

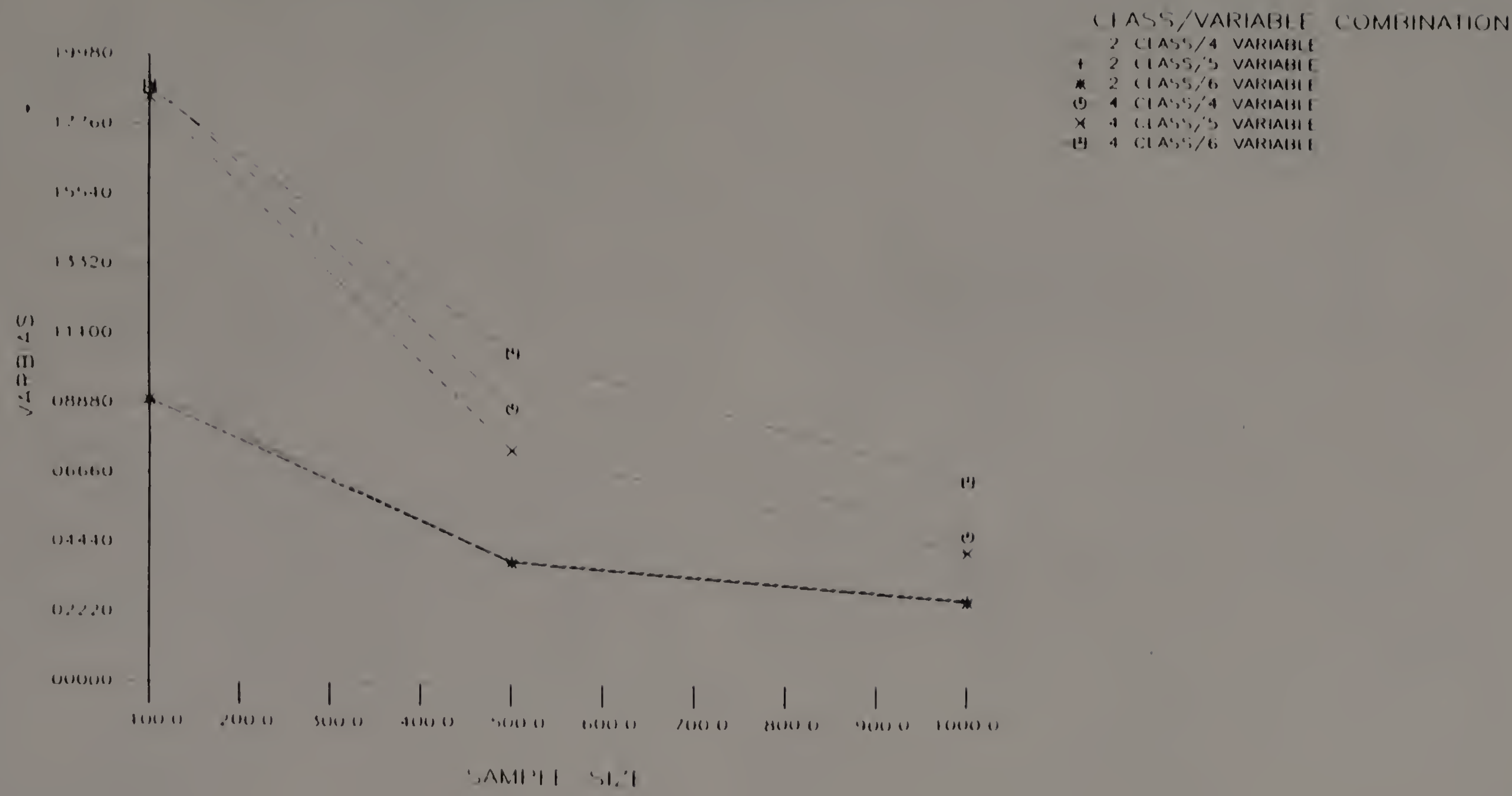

FIGURE 4.16 -- VARBIAS BY NVAR*NCLASS*DISC

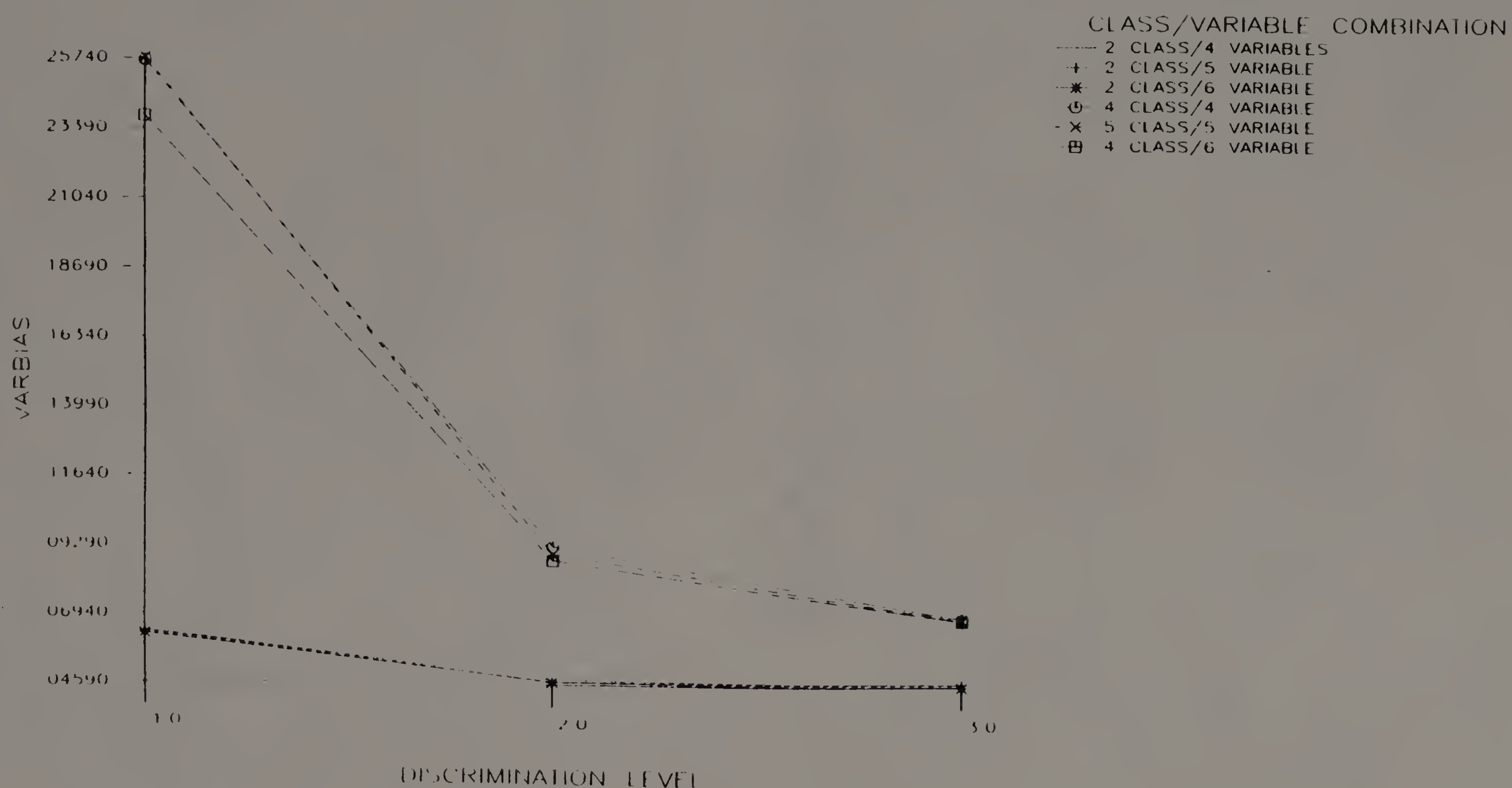

FIGURE 4.17 VARBIAS BY NCLASS*NGROUP*DISC

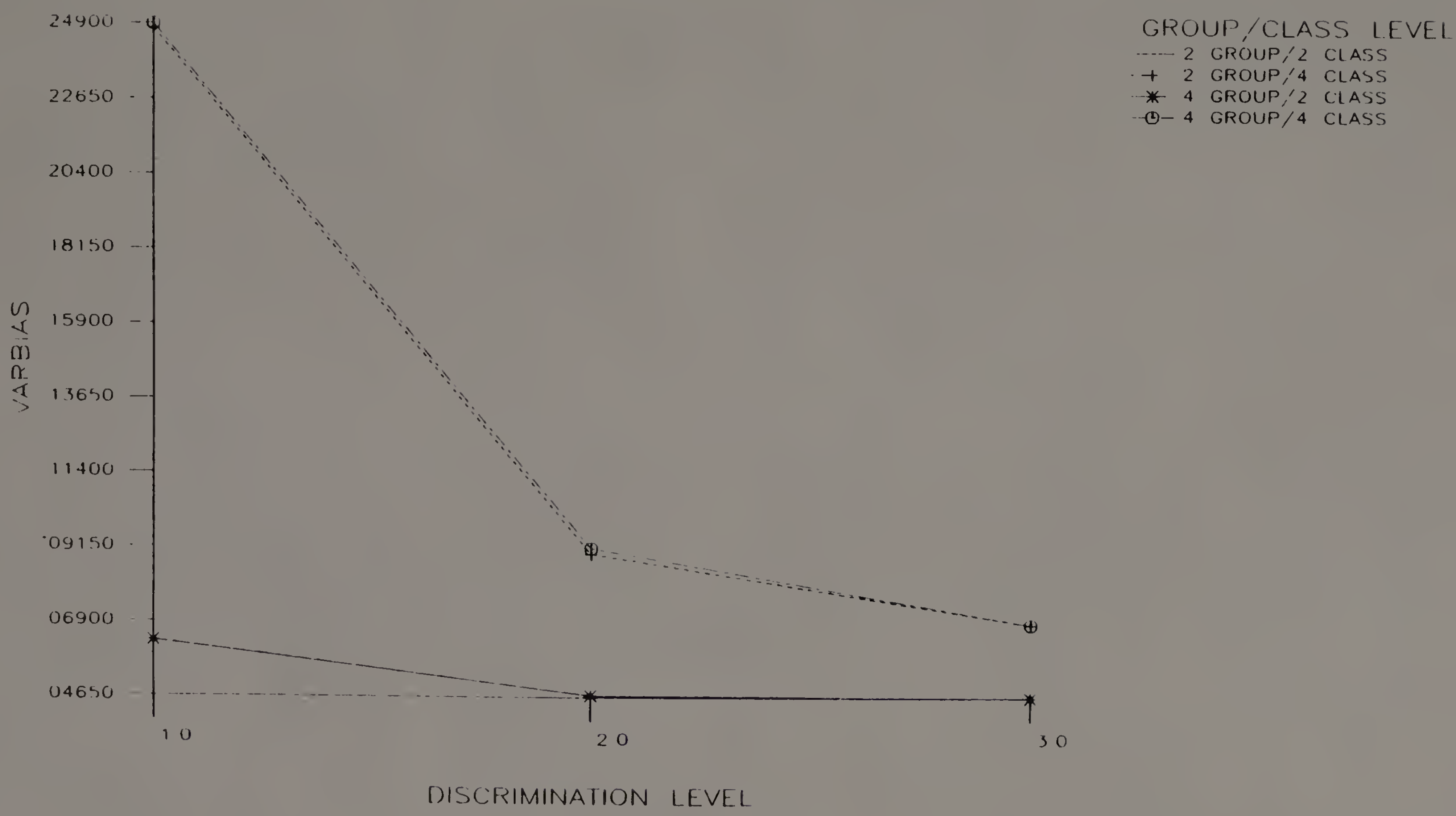

FIGURE 4.18 - VARBIAS BY NVAR*NCASE*DISC

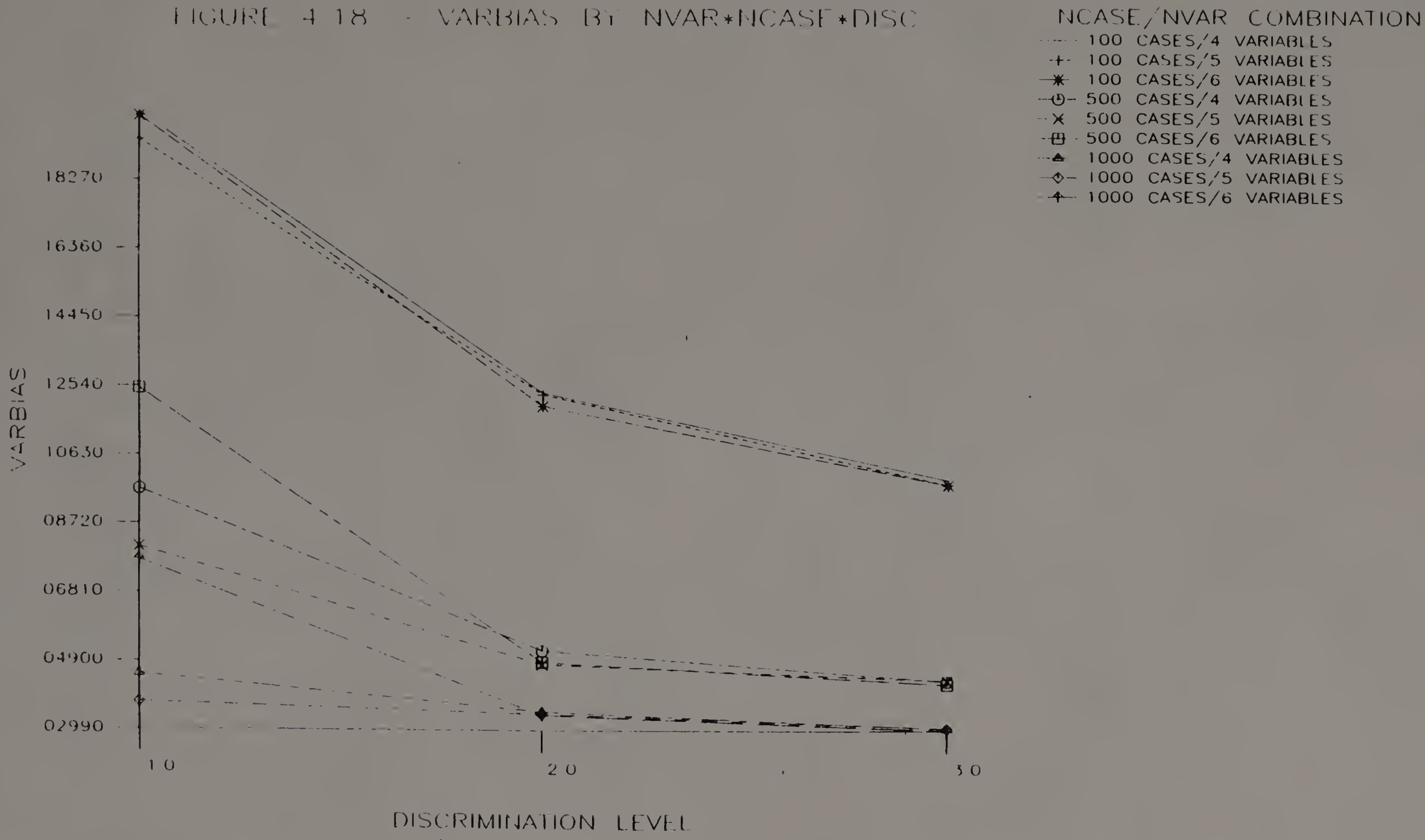

FIGURE 4.19 VARBIAS BY NCLASS*NCASE*DISC

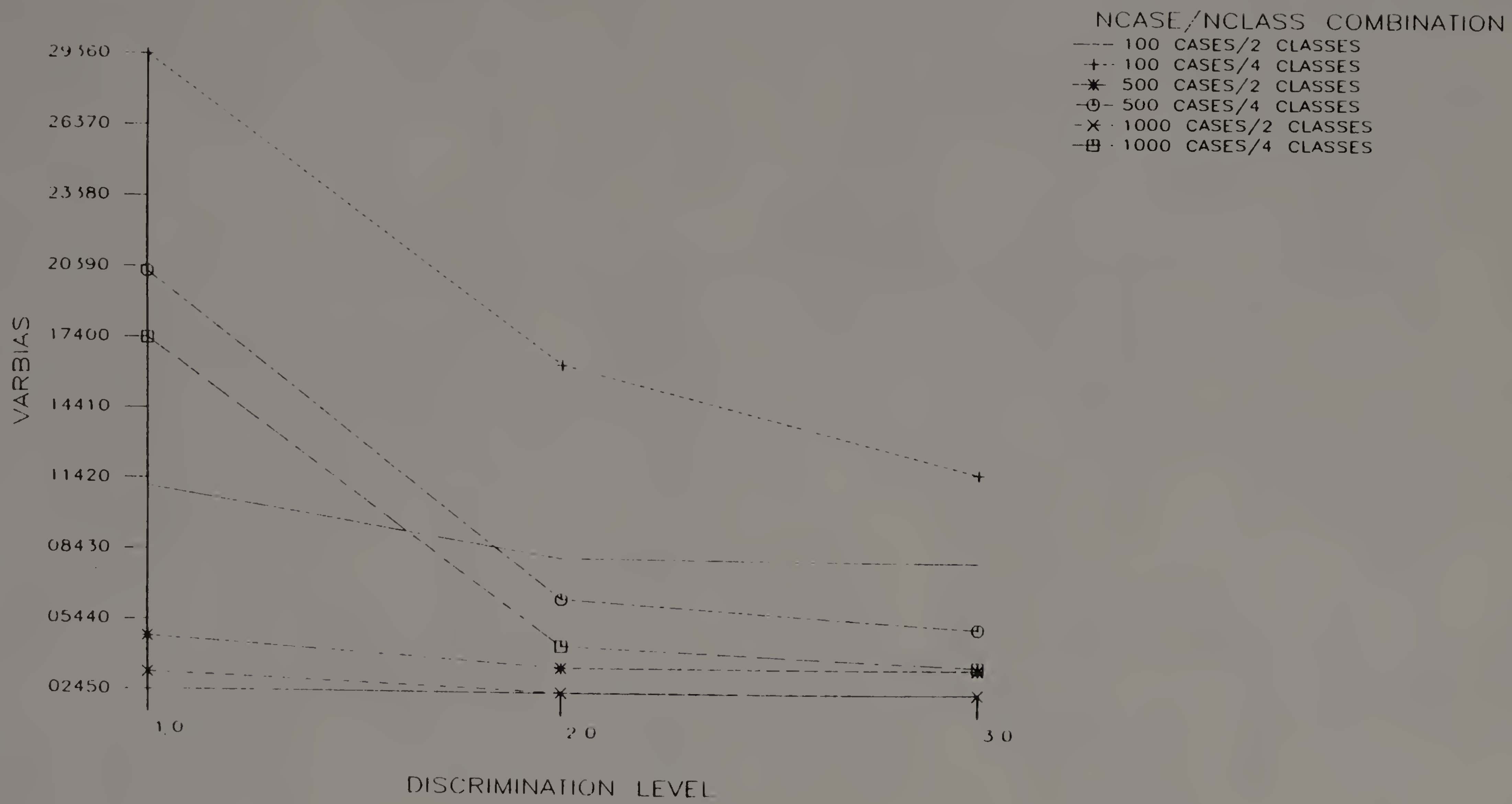

FIGURE 4.20 CONBIAS BY NVAR*NCLASS*NCASE

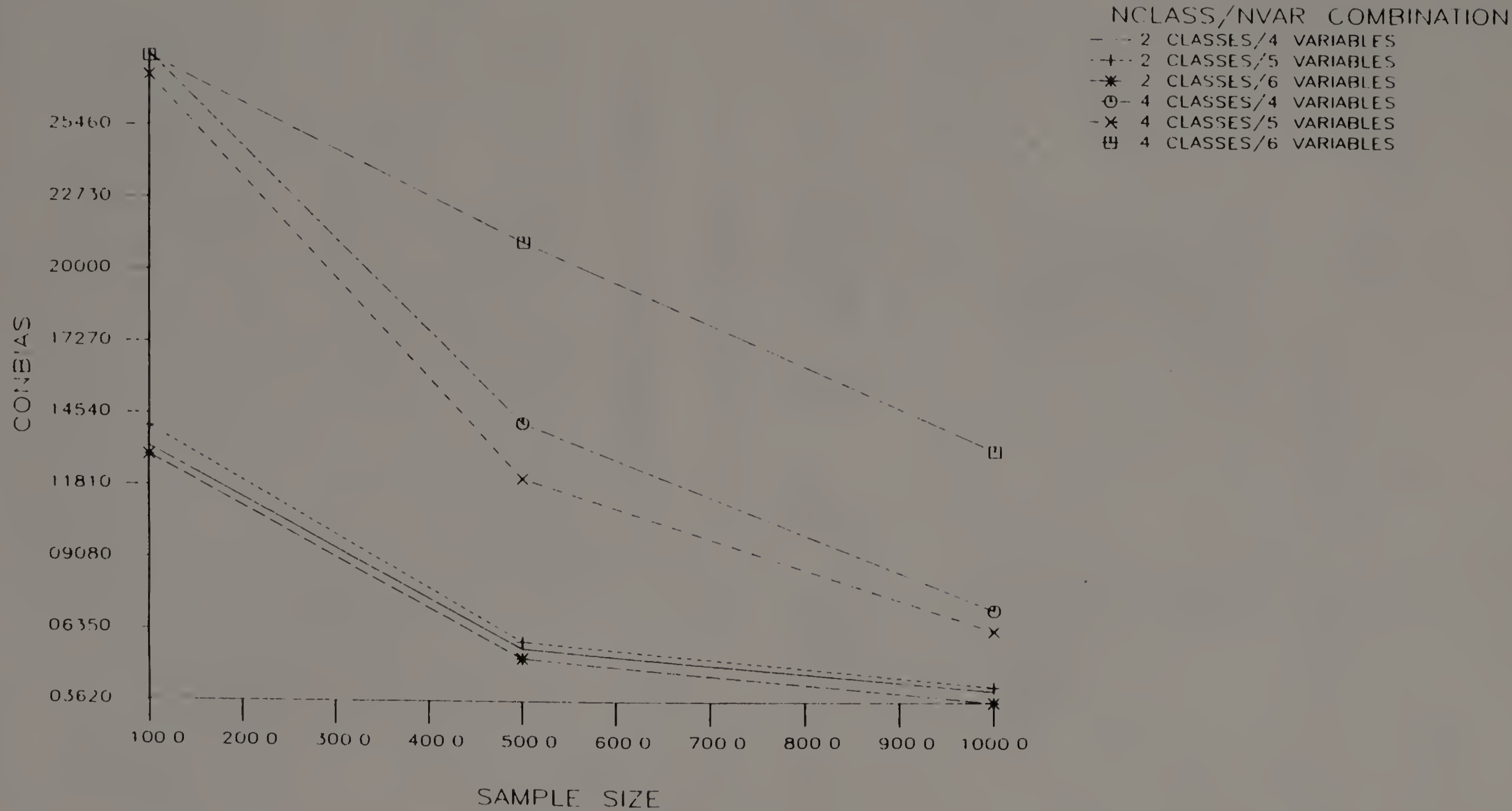

FIGURE 4.21 -- CONBIAS BY NVAR*NCLASS*DISC

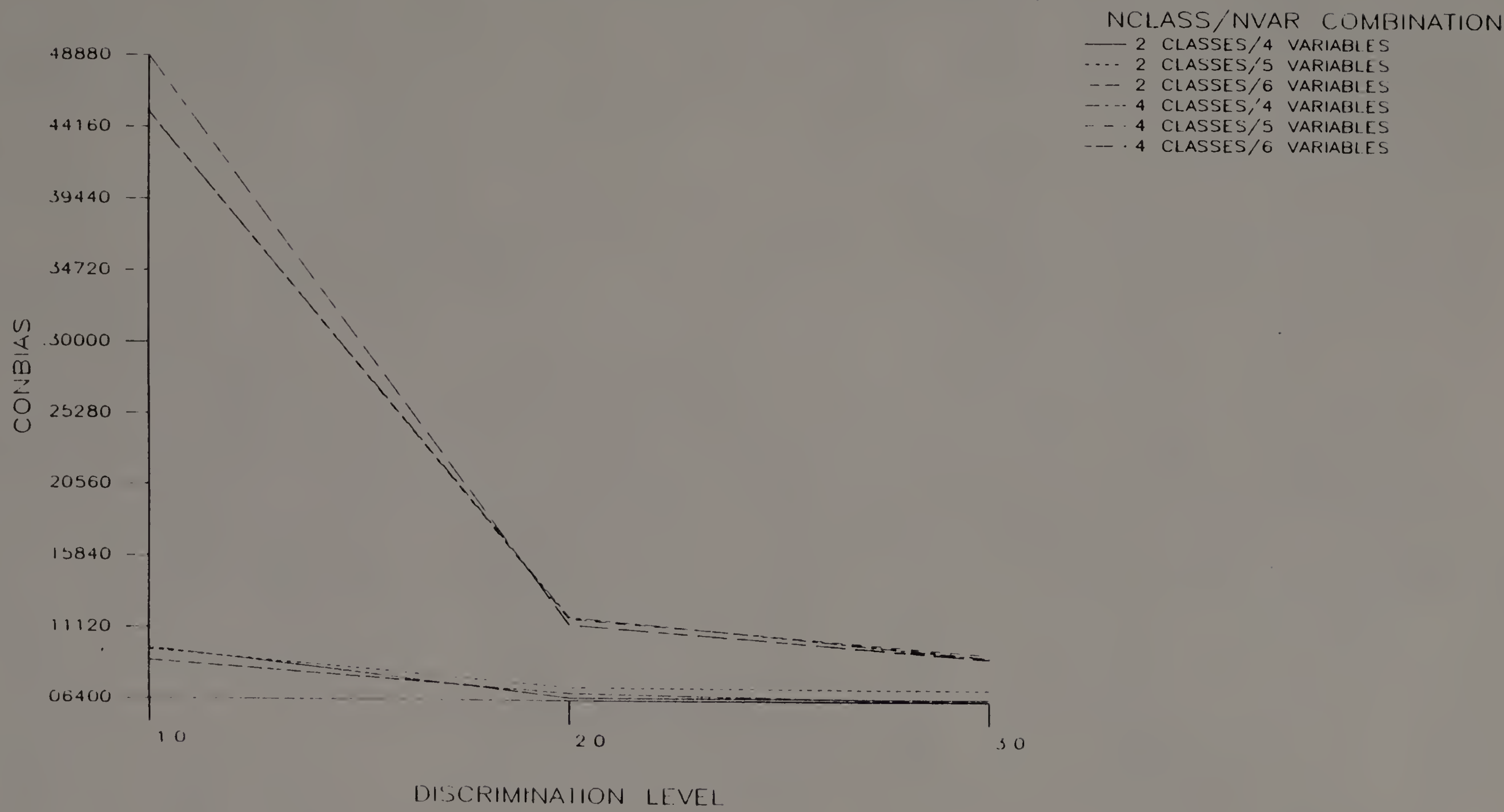

FIGURE 4.22 - CONBIAS BY NCLASS*NGROUP*DISC

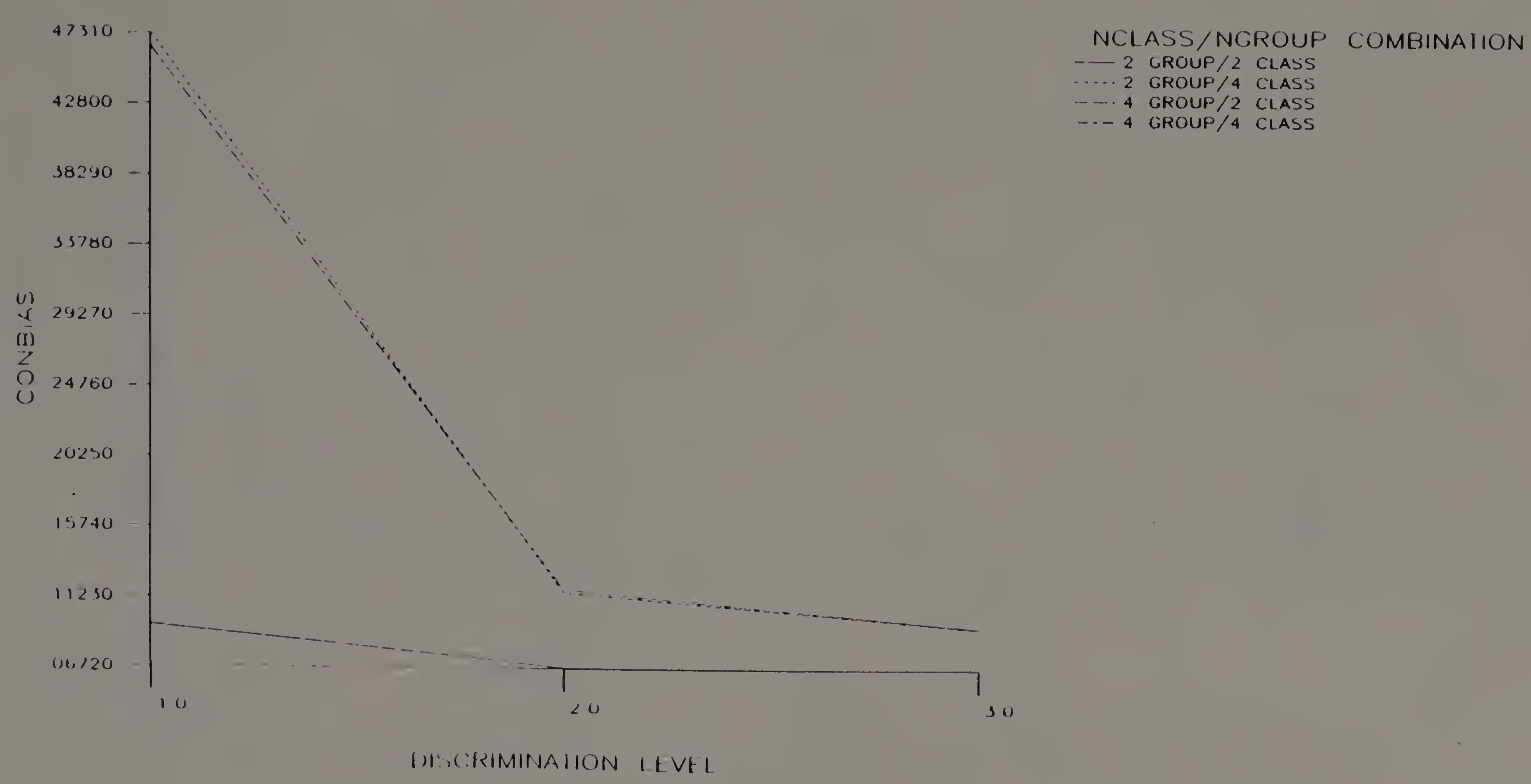

FIGURE 4.23 CONBIAS BY NVAR*NCASE*DISC

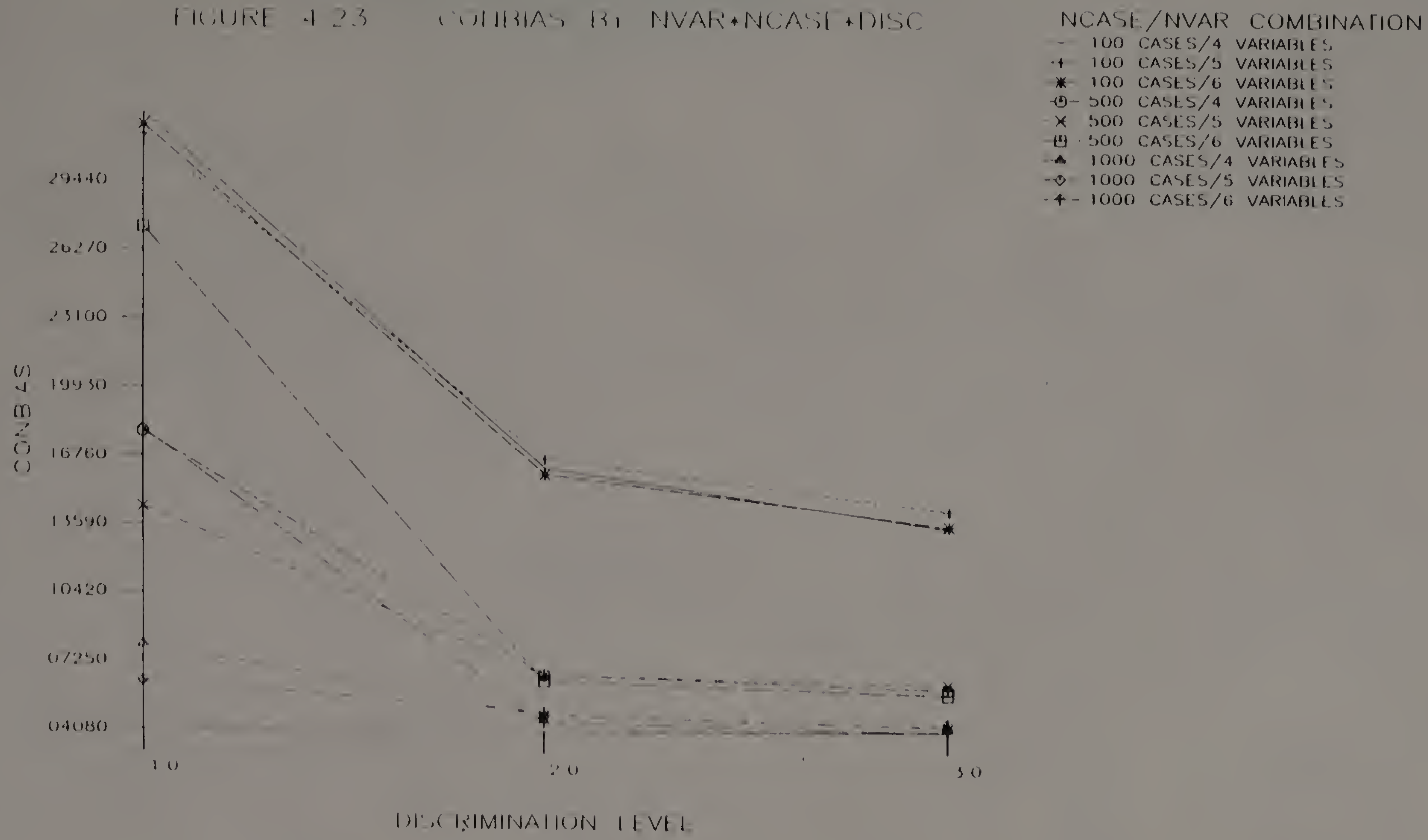

FIGURE 4.24 — CONBIAS BY NCLASS*NCASE*DISC

FIGURE 4.25 CATBIAS BY NVAR*NCLASS*NCASE

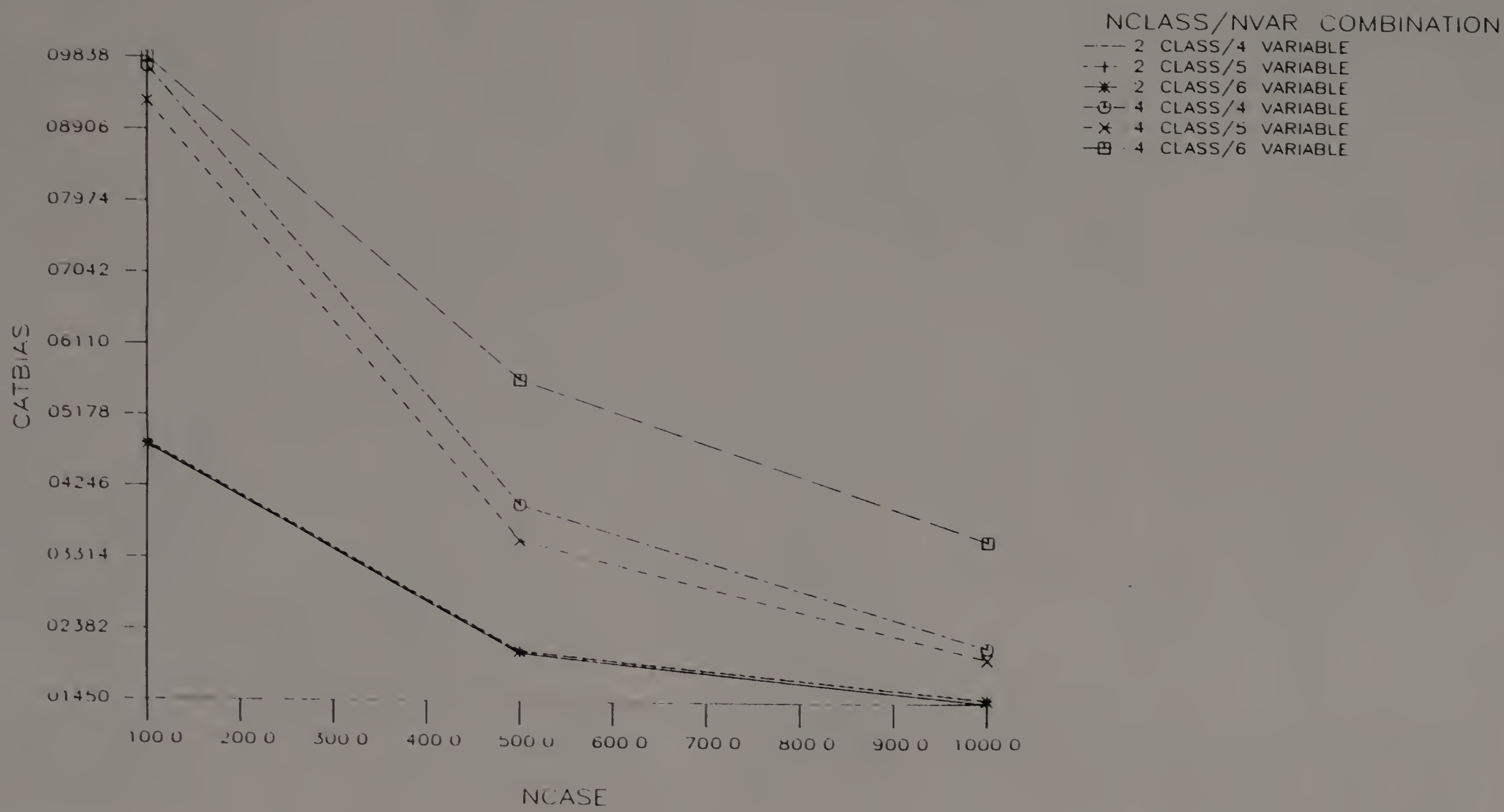

FIGURE 4.26 - CATBIAS BY NVAR*NCLASS*DISC

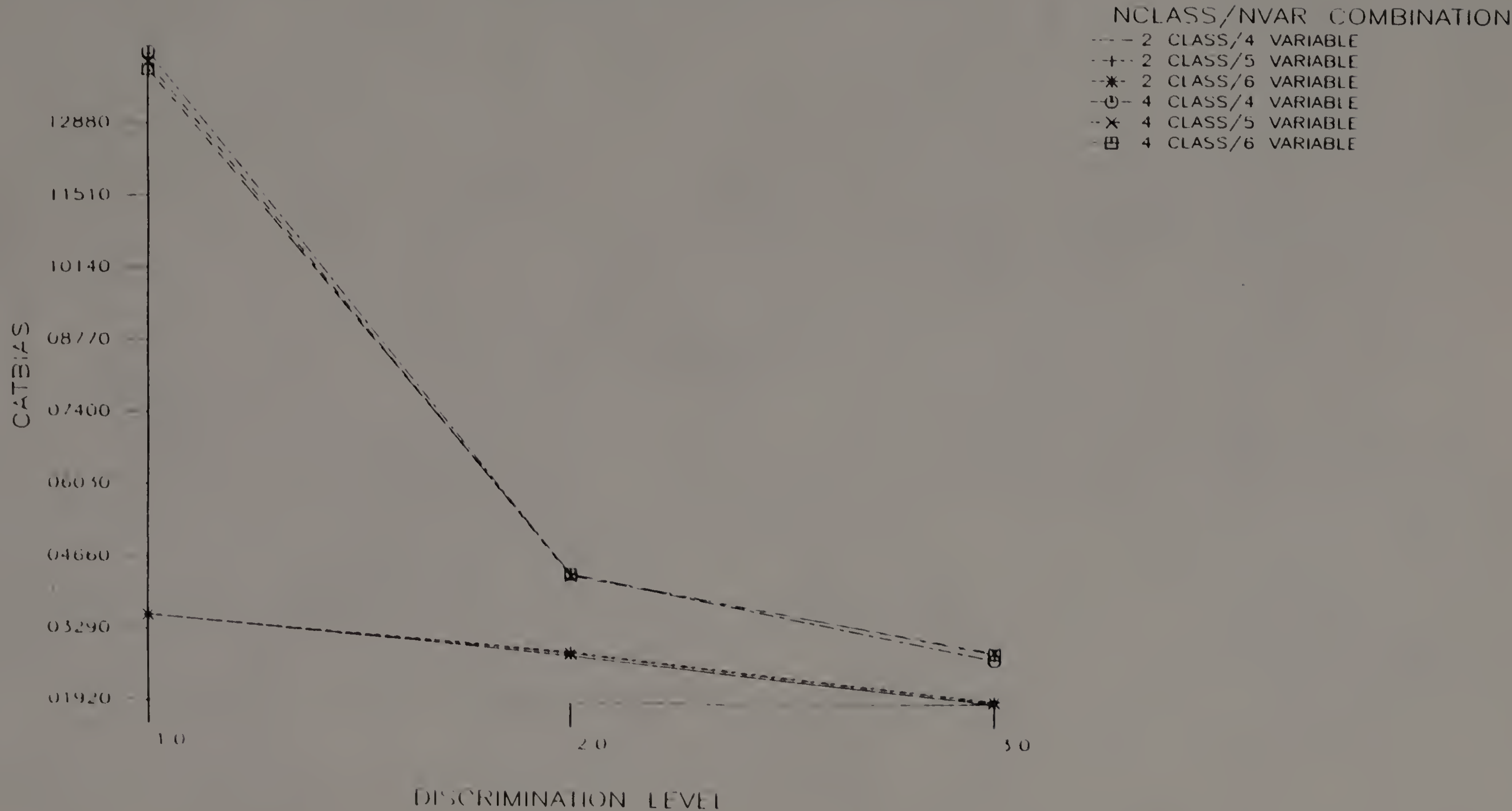

FIGURE 4.27 CATBIAS BY NVAR*NCASE*DISC

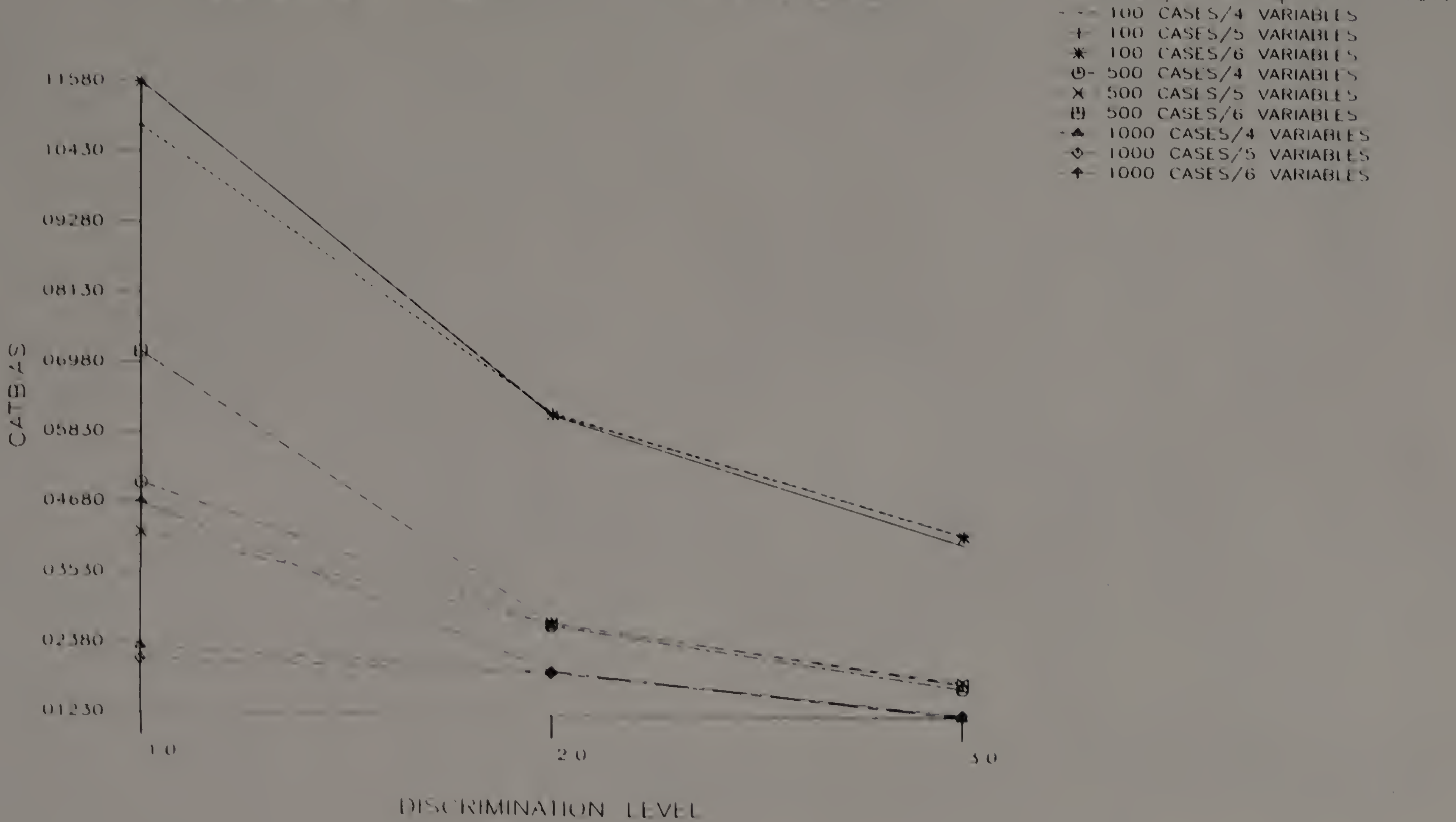

FIGURE 4.28 CATBIAS BY NCLASS*NCASE*DISC

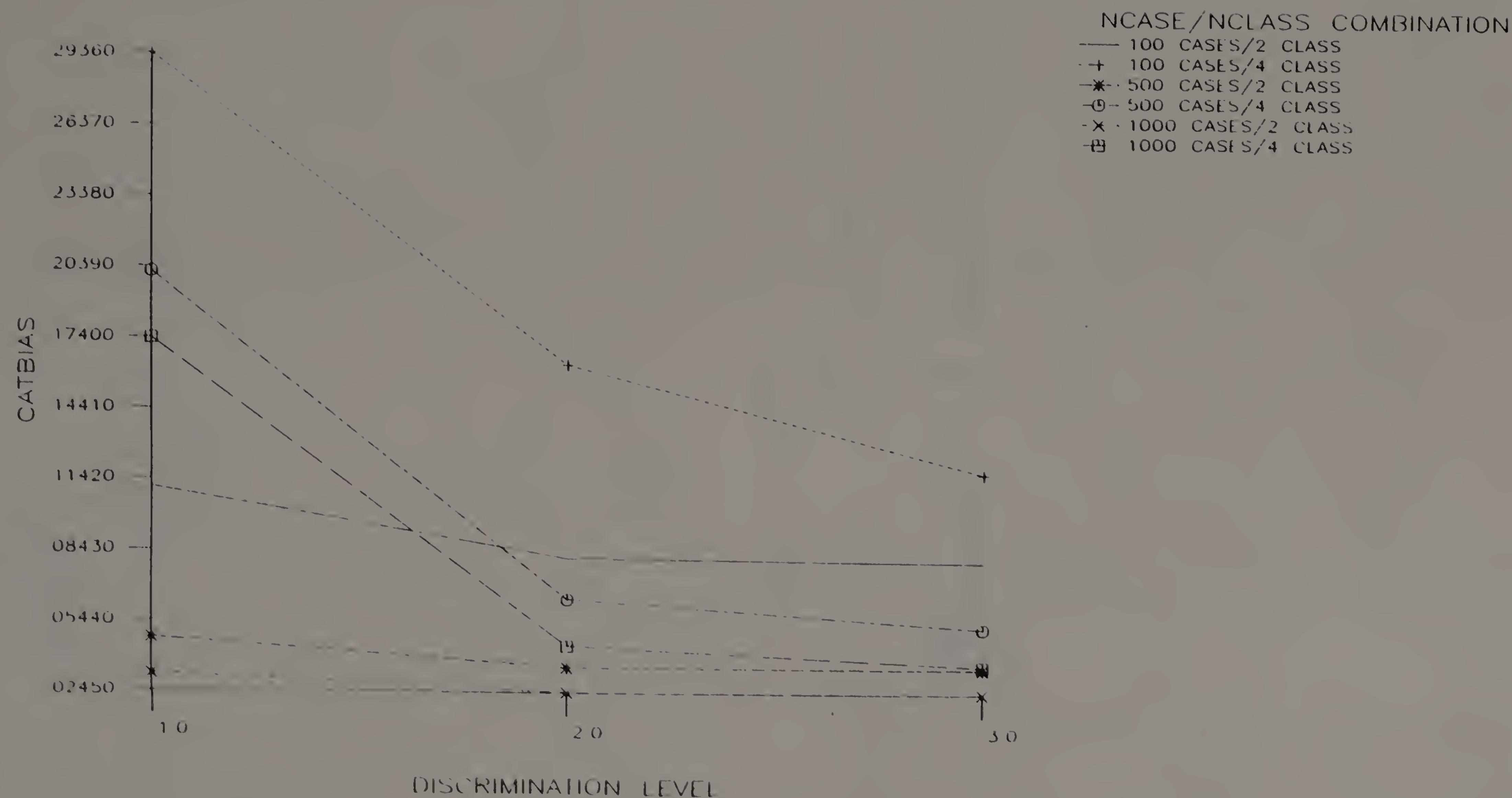

FIGURE 4.29 - GBIAS BY NVAR*NCLASS*NCASE

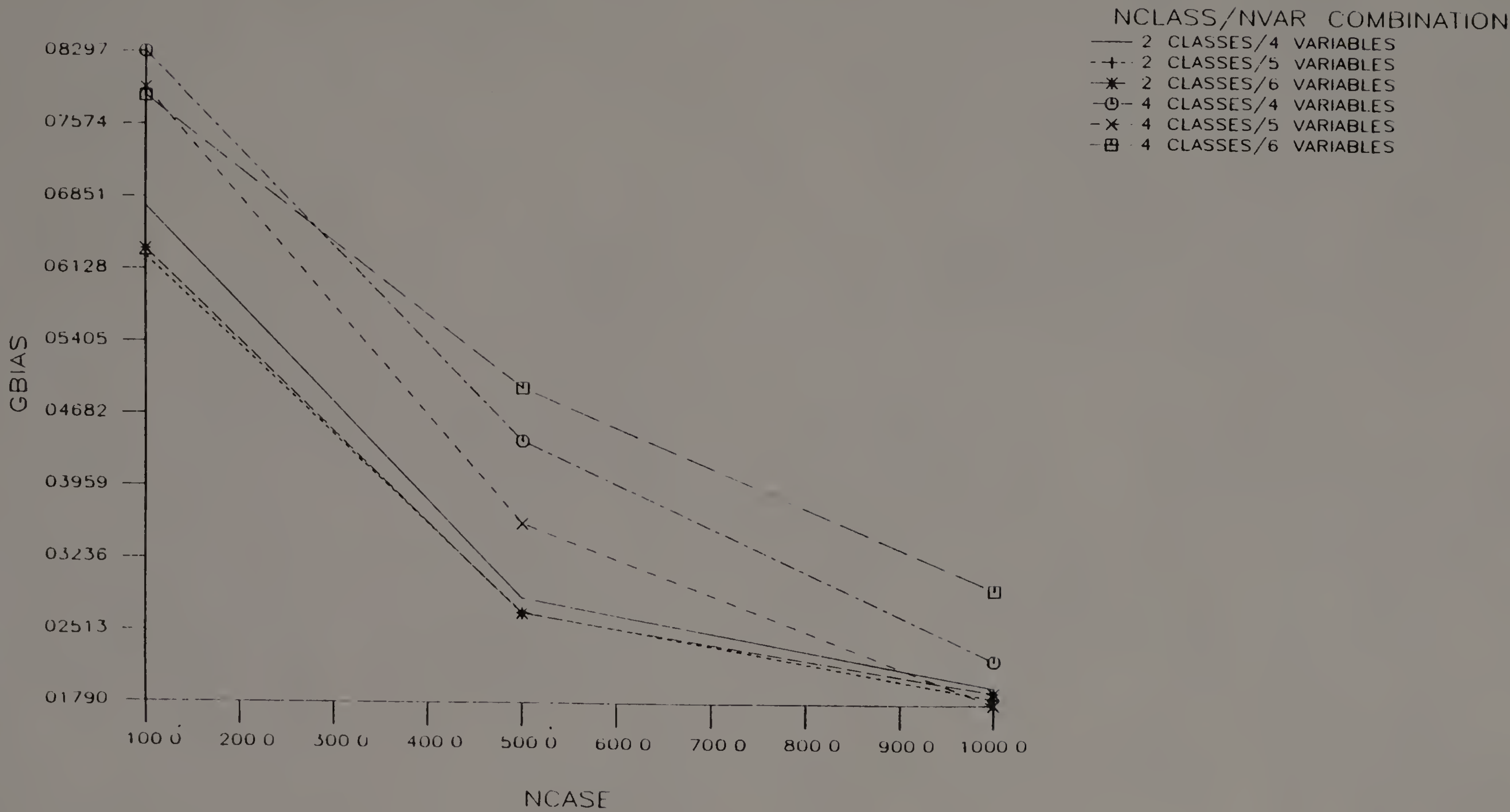

FIGURE 4.30 - GBIAS BY NVAR*NCLASS*DISC

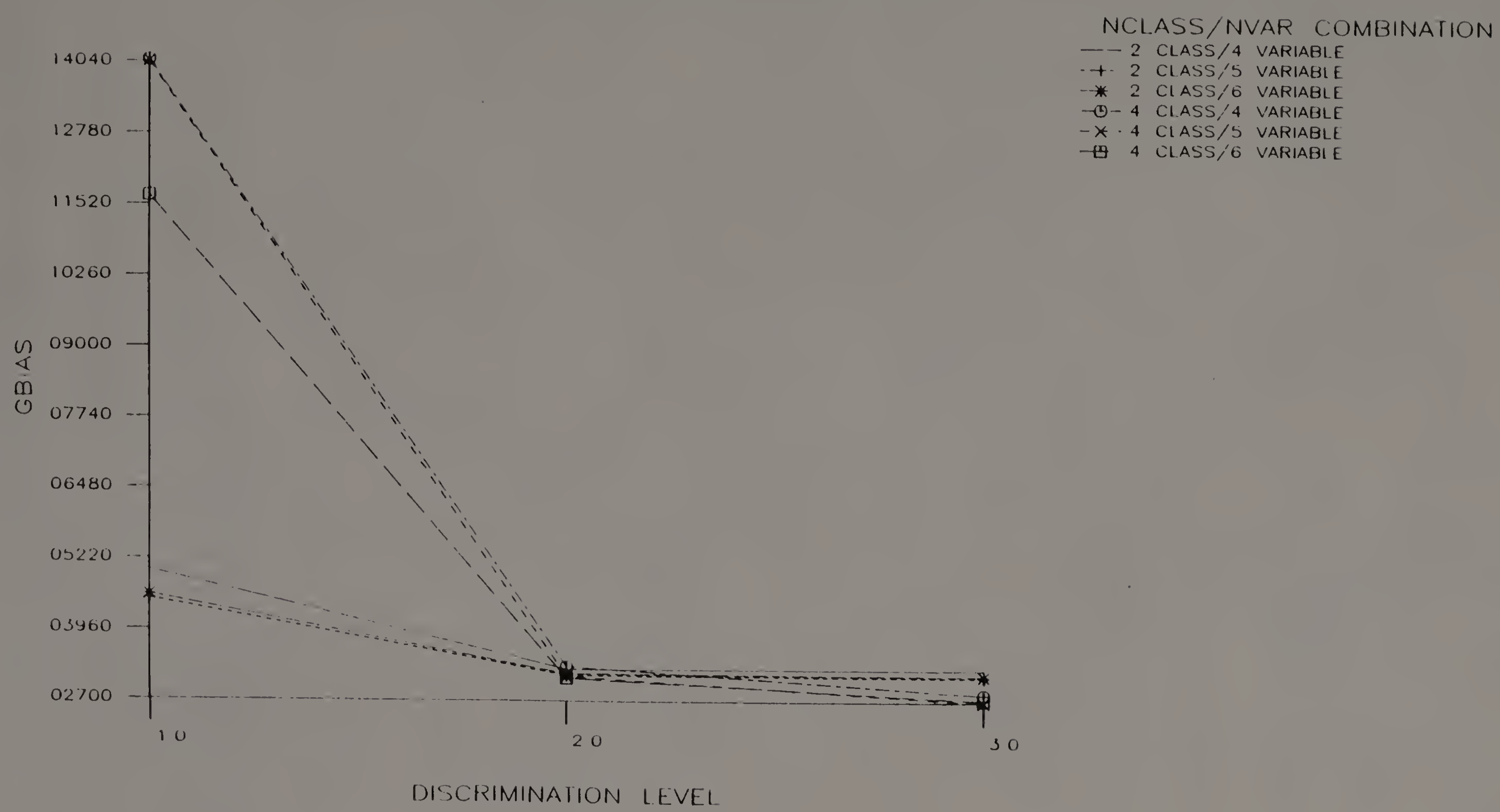

FIGURE 4.31 GBIAS BY NVAR*NGROUP*NCASE

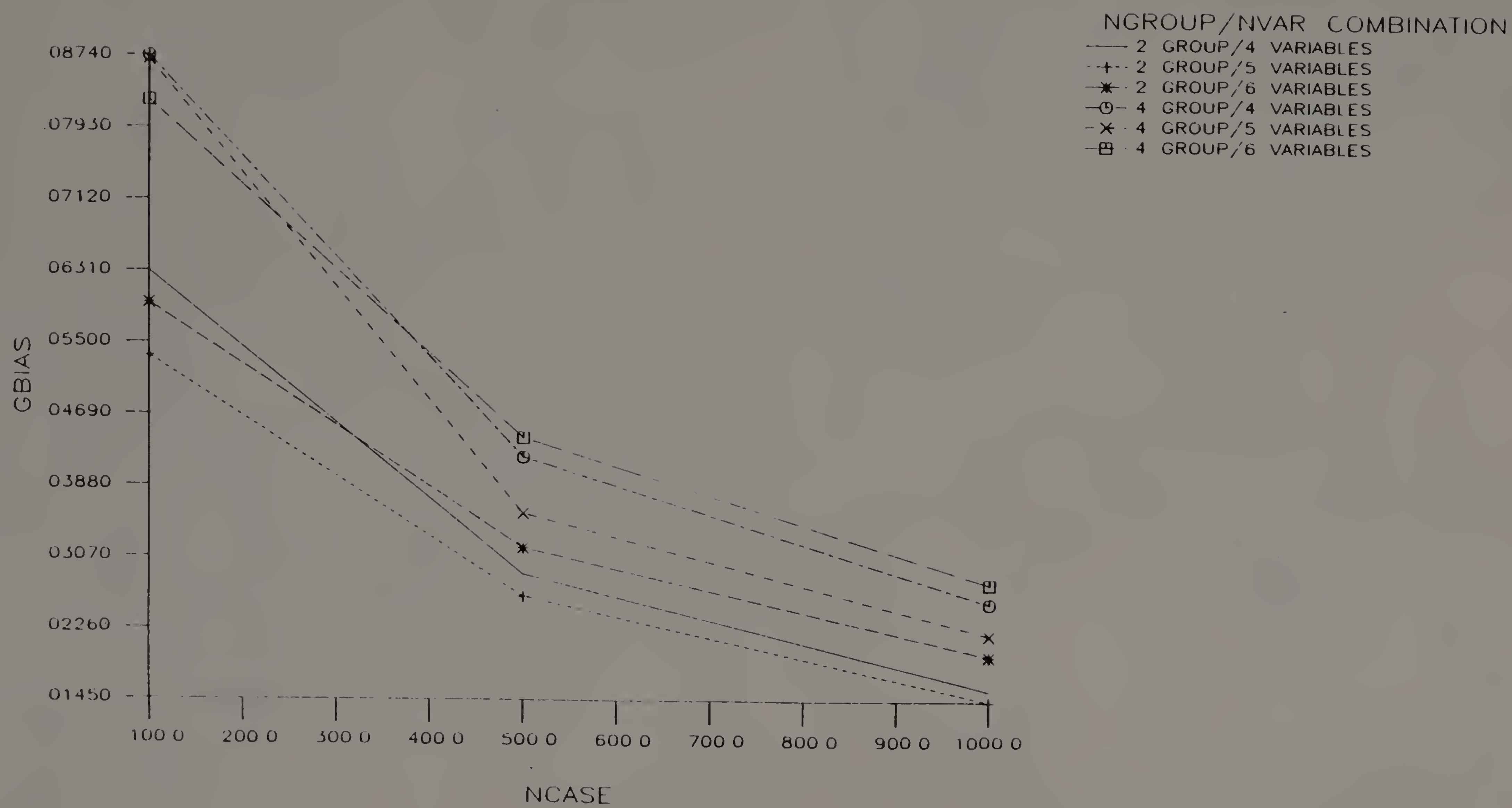

FIGURE 4.32 - GBIAS BY NCLASS*NGROUP*DISC

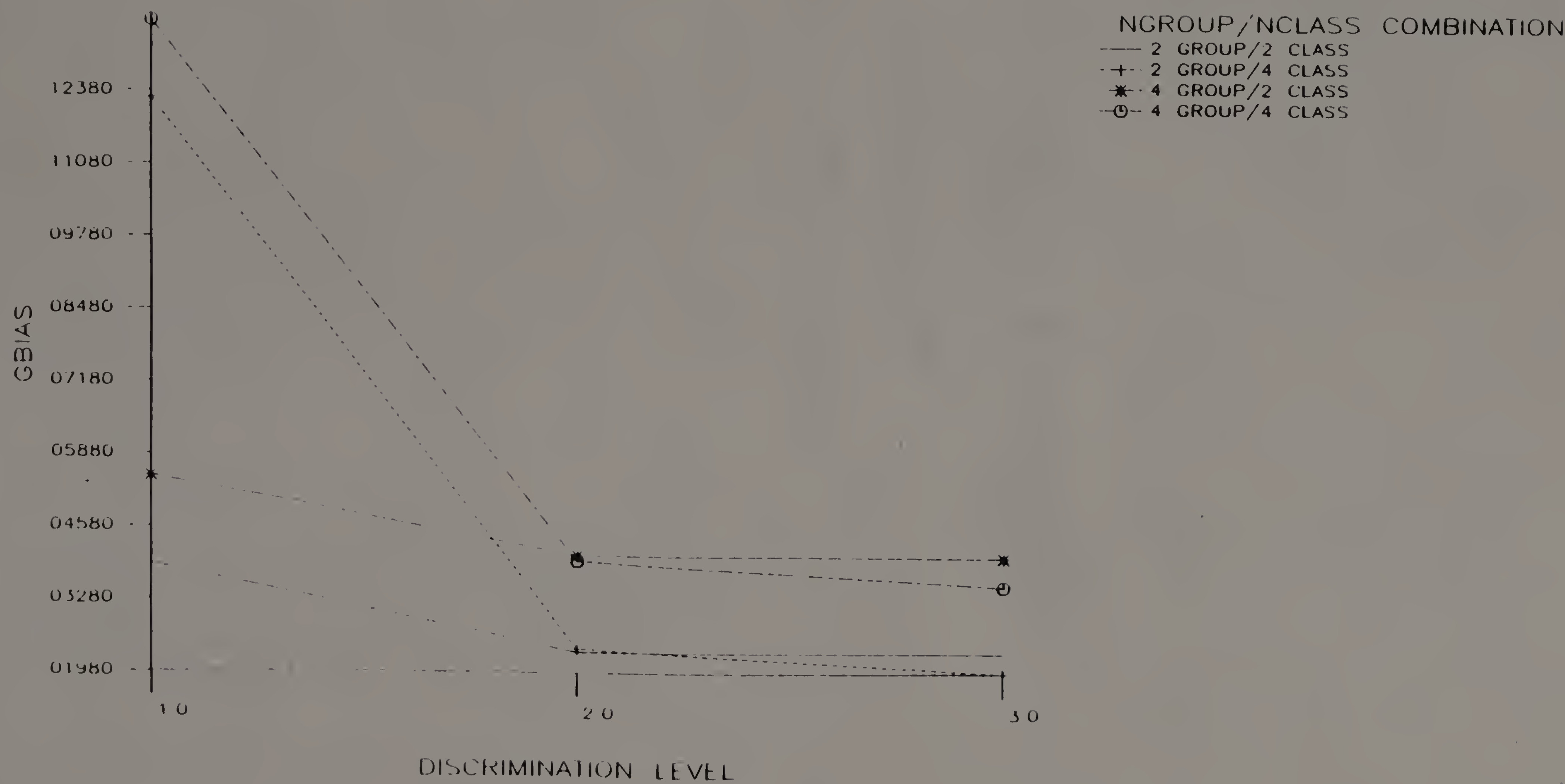

FIGURE 4.33 - GBIAS BY NCLASS*NCASE*DISC

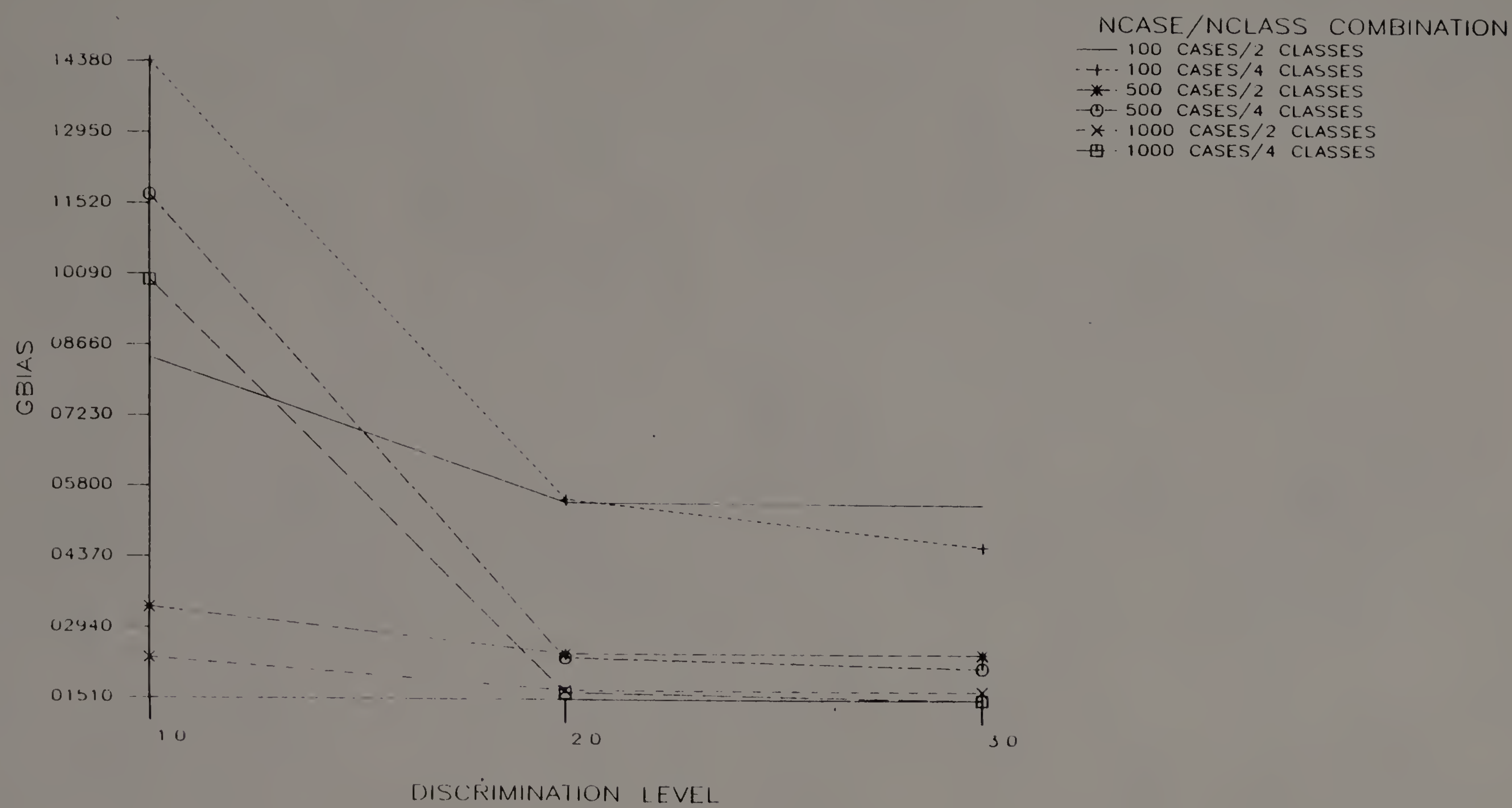

FIGURE 4.34 ZKBIAS BY NVAR*NCLASS*DISC

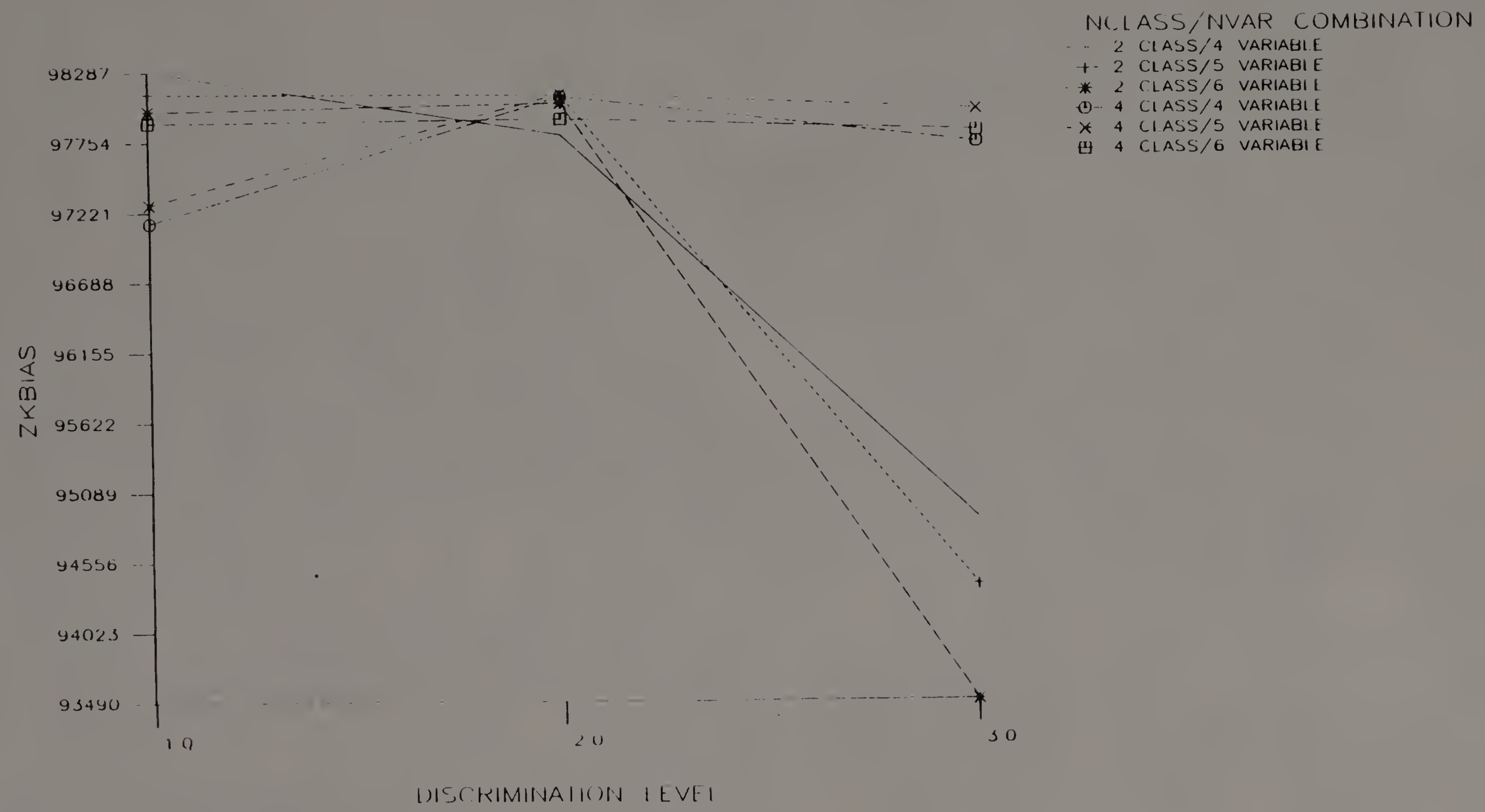

FIGURE 4.35 ZKBIAS BY NVAR*NCASE*DISC

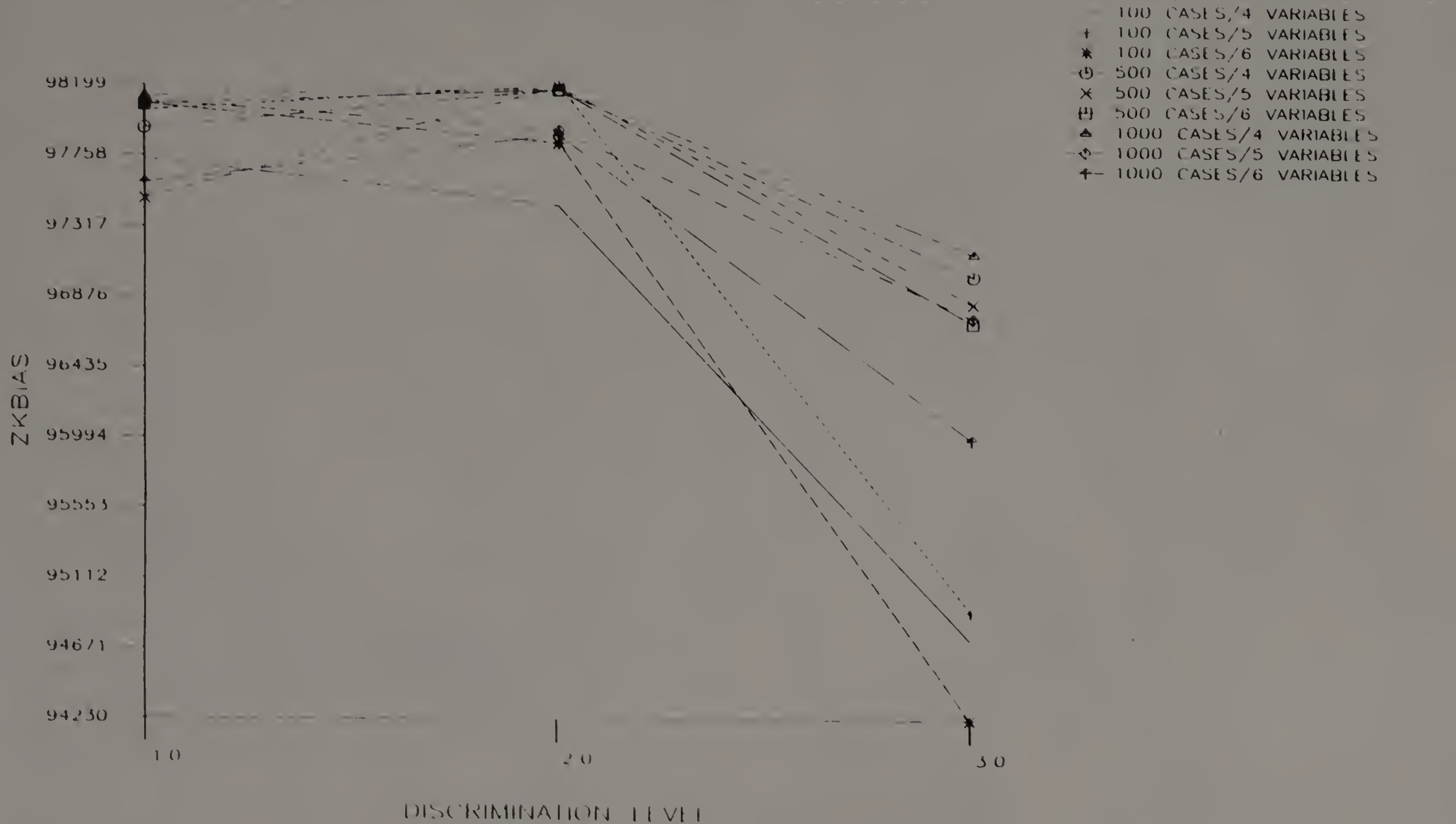

FIGURE 4.36 — TOTBIAS BY NVAR*NCLASS*NGROUP

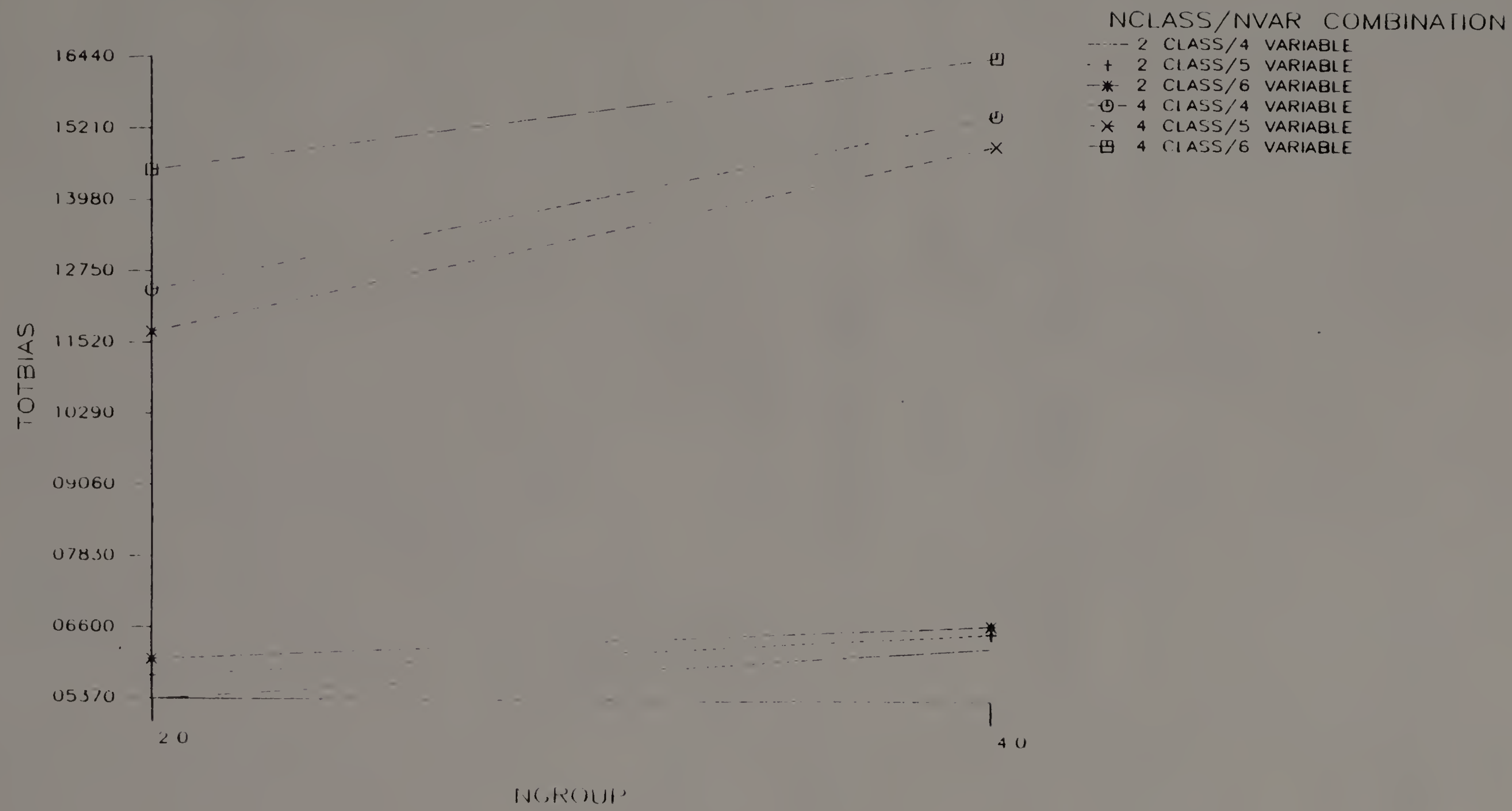

FIGURE 4.37 – TOTBIAS BY NVAR*NCLASS*NCASE

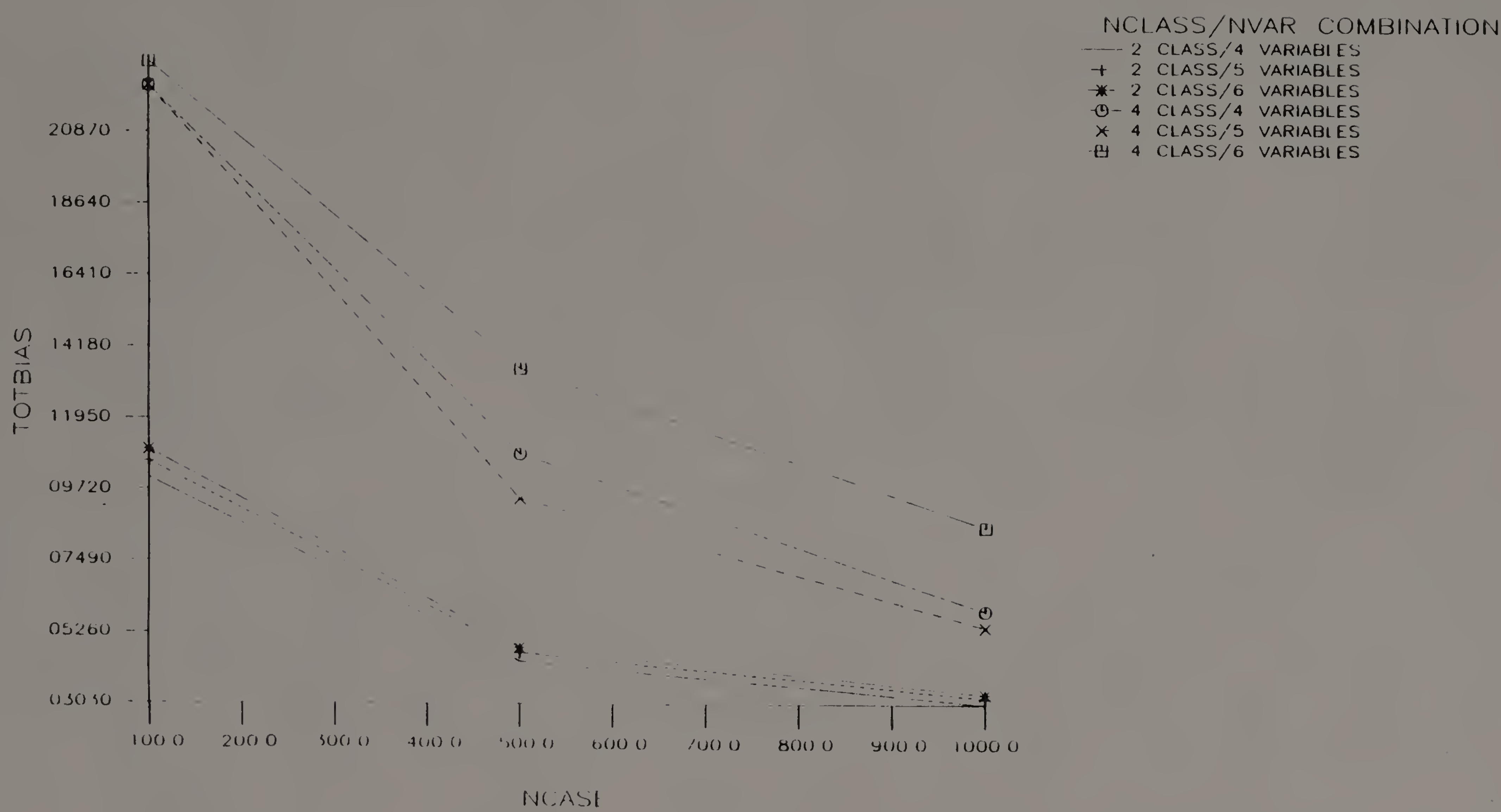

FIGURE 4.38 TOTBIAS BY NVAR*NCLASS*DISC

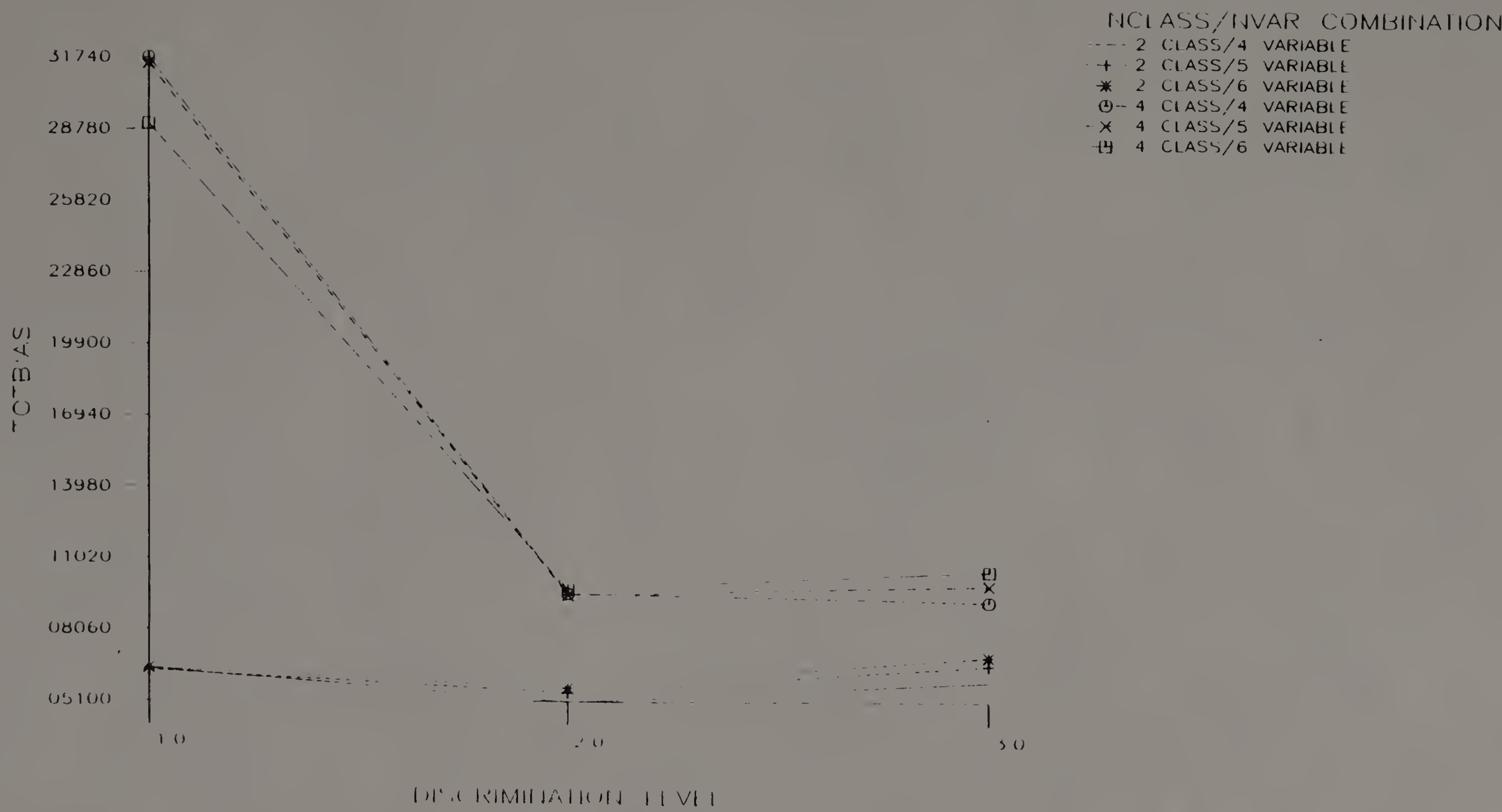

FIGURE 4.39 TOTBIAS BY NVAR*NGROUP*NCASE

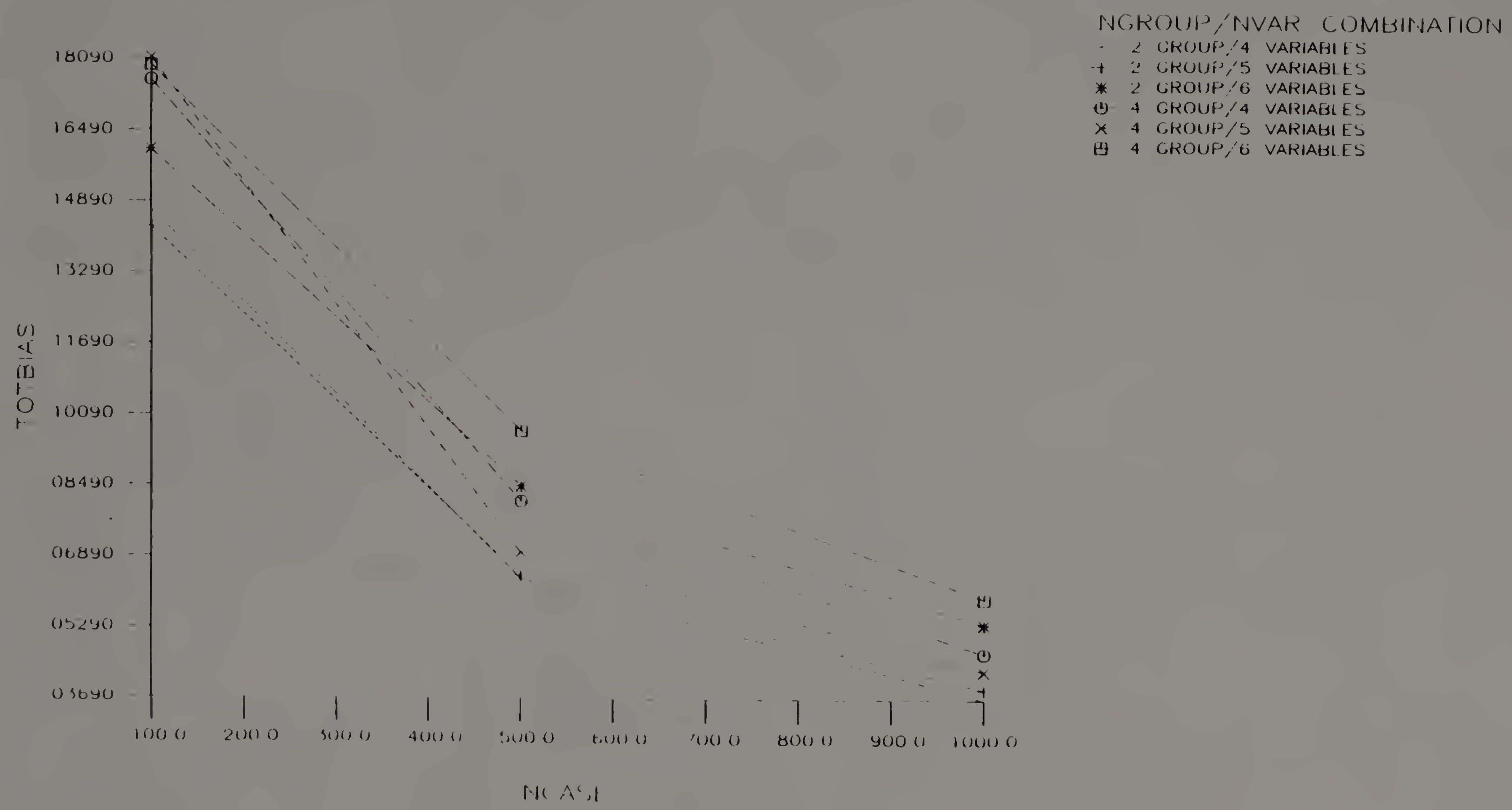

FIGURE 4.40 - TOTBIAS BY NCLASS*NGROUP*DISC

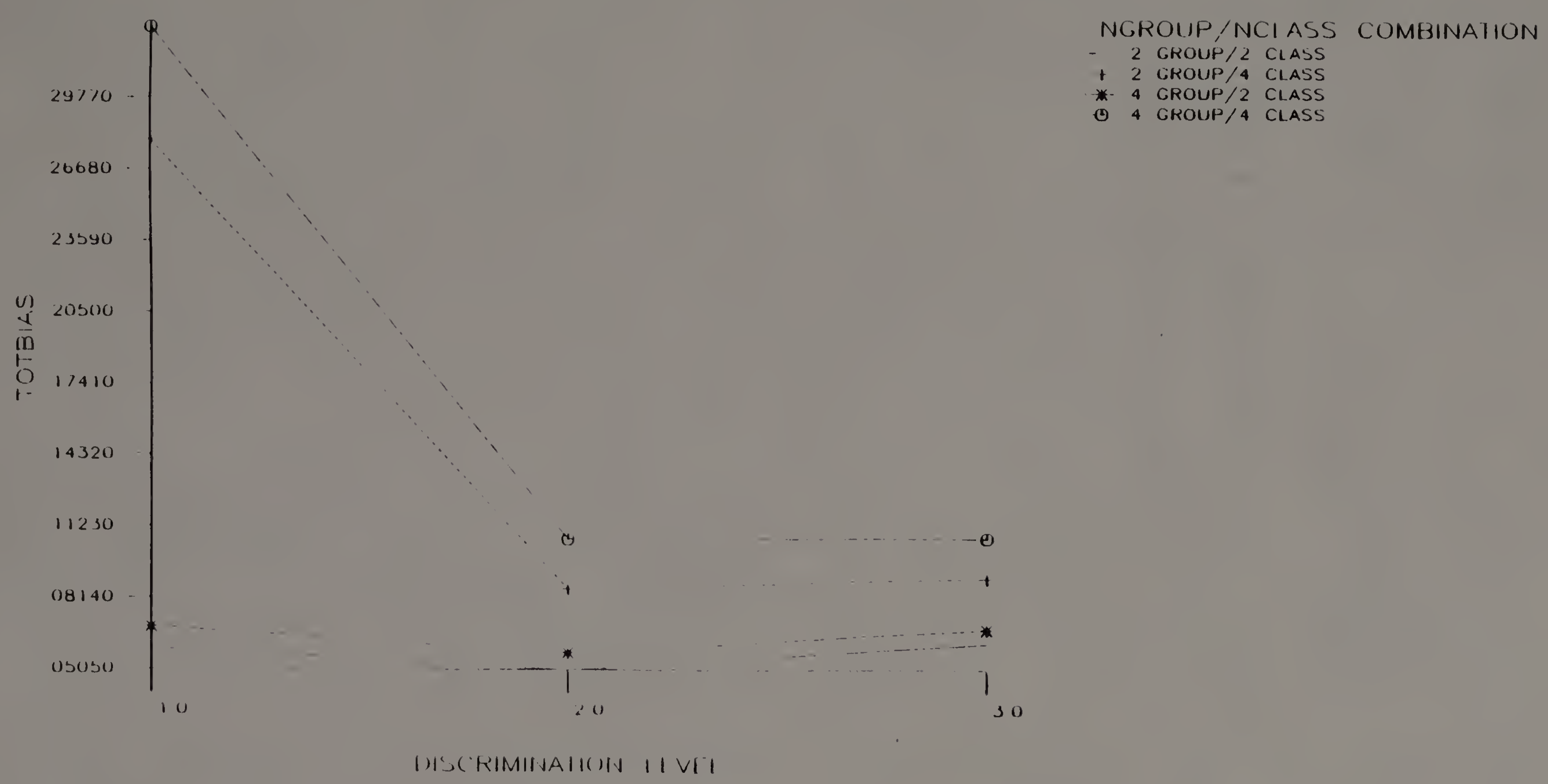

FIGURE 4.41 - TOTBIAS BY NVAR*NCASE*DISC

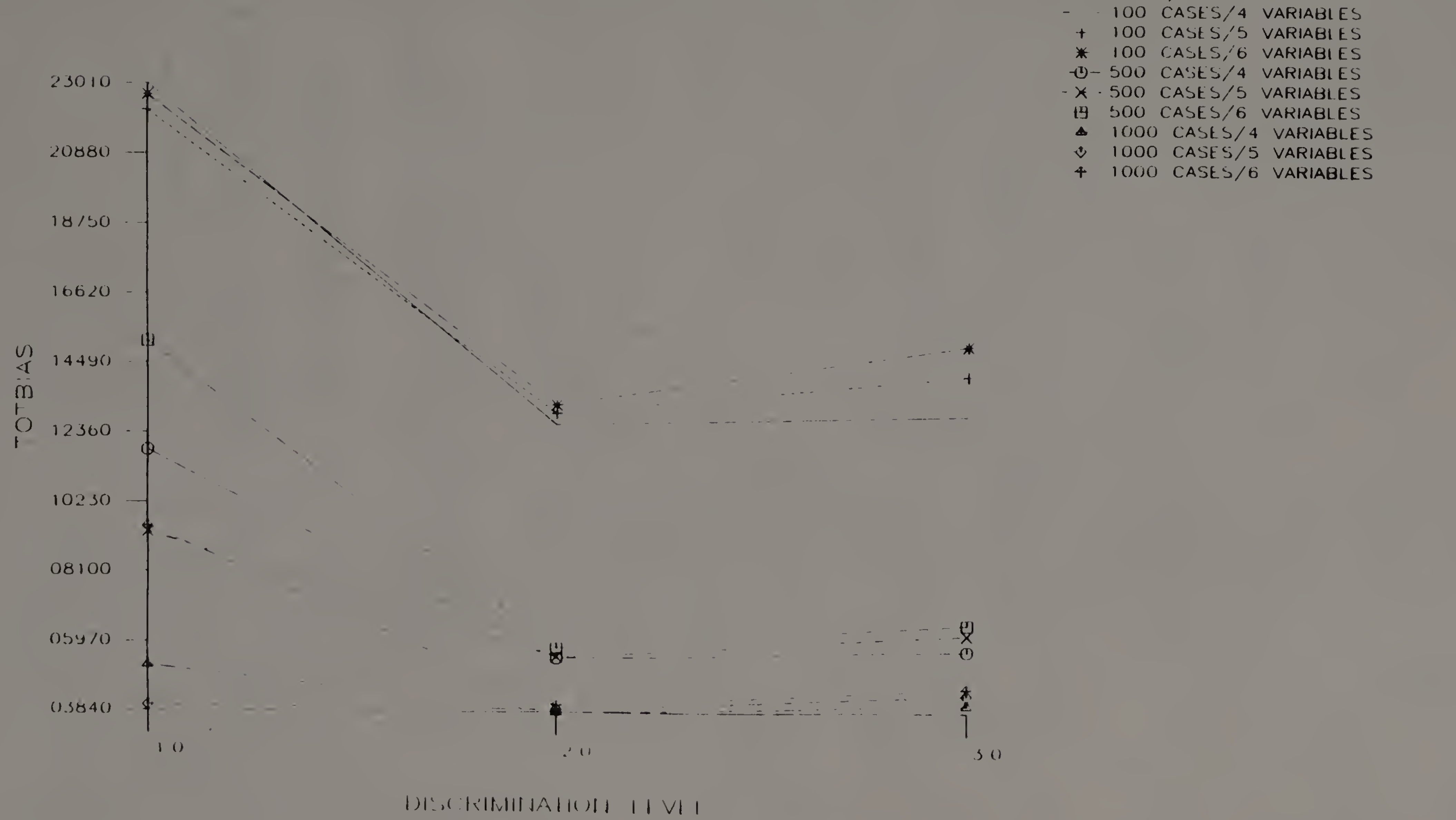

FIGURE 4.42 TOTBIAS BY NCLASS*NCASE*DISC

FIGURE 4.43 - TOTBIAS BY NGROUP*NCASE*DISC

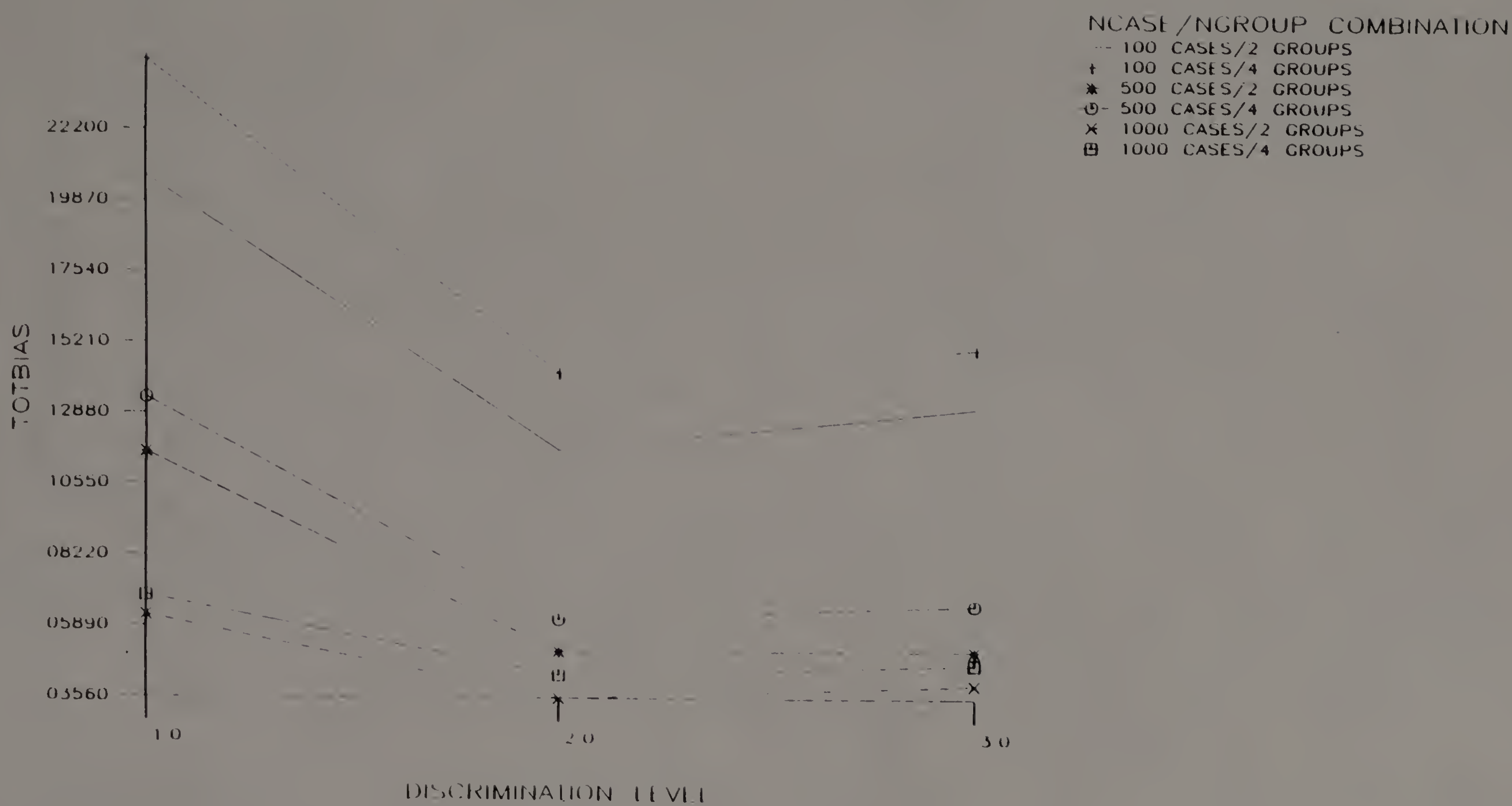

FIGURE 4.44 — TOTBIAS BY NCLASS*NGROUP*NCASE

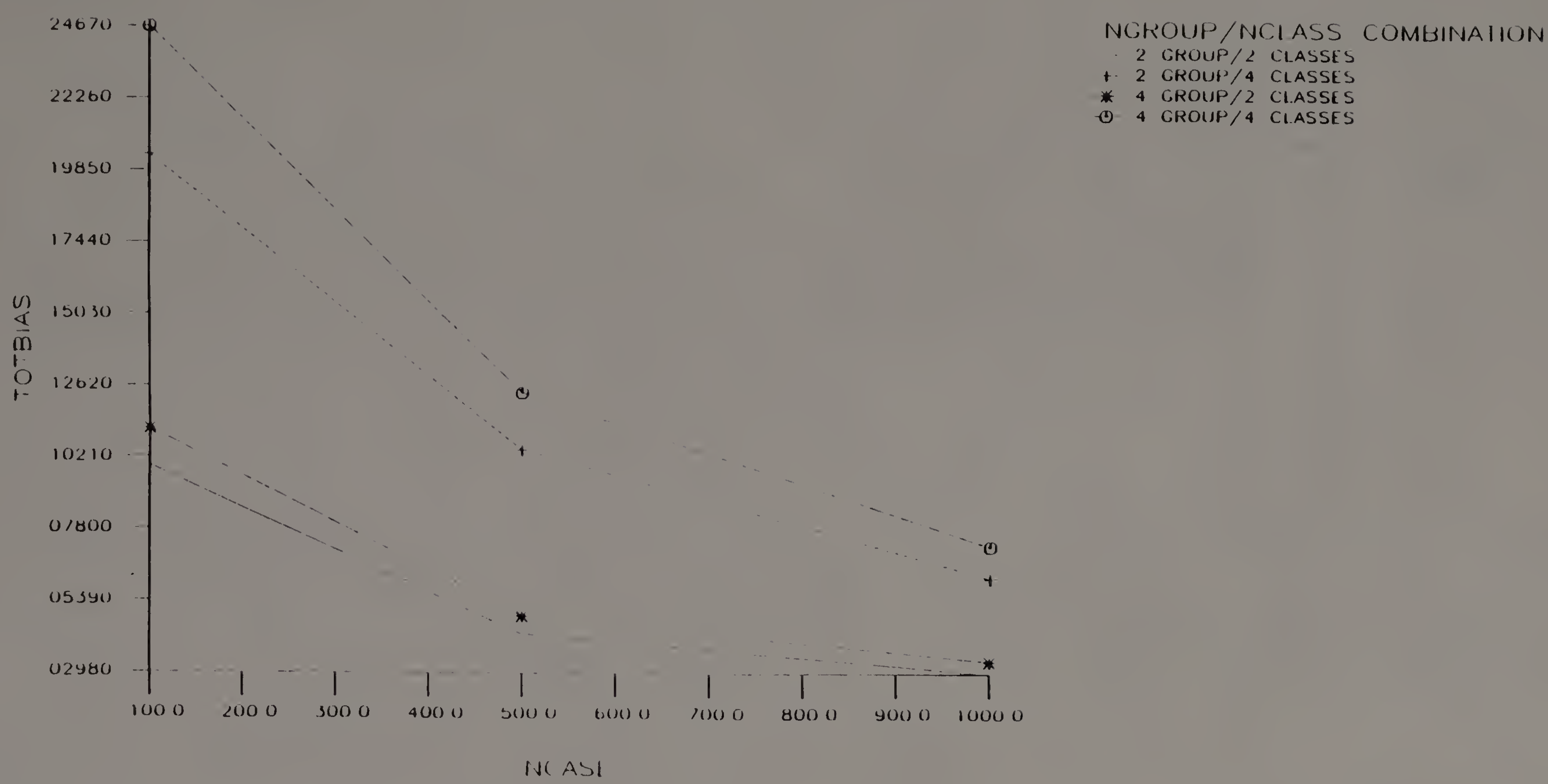

TABLE 1

SIMULATION DESIGN

Cell No.	No. of Variables	No. of Classes	No. of Groups	Discrimination* Level	Sample Size
1	4	2	2	L	100
2	4	2	2	L	500
3	4	2	2	L	1000
4	4	2	2	M	100
5	4	2	2	M	500
6	4	2	2	M	1000
7	4	2	2	W	100
8	4	2	2	W	500
9	4	2	2	W	1000
10	4	2	4	L	100
11	4	2	4	L	500
12	4	2	4	L	1000
13	4	2	4	M	100
14	4	2	4	M	500
15	4	2	4	M	1000
16	4	2	4	W	100
17	4	2	4	W	500
18	4	2	4	W	1000
19	4	4	2	L	100
20	4	4	2	L	500
21	4	4	2	L	1000
22	4	4	2	M	100
23	4	4	2	M	500
24	4	4	2	M	1000
25	4	4	2	W	100
26	4	4	2	W	500
27	4	4	2	W	1000
28	4	4	4	L	100
29	4	4	4	L	500
30	4	4	4	L	1000
31	4	4	4	M	100
32	4	4	4	M	500
33	4	4	4	M	1000
34	4	4	4	W	100
35	4	4	4	W	500
36	4	4	4	W	1000
37	5	2	2	L	100
38	5	2	2	L	500
39	5	2	2	L	1000
40	5	2	2	M	100

TABLE 1
(continued)

Cell No.	No. of Variables	No. of Classes	No. of Groups	Discrimination* Level	Sample Size
41	5	2	2	M	500
42	5	2	2	M	1000
43	5	2	2	W	100
44	5	2	2	W	500
45	5	2	2	W	1000
46	5	2	4	L	100
47	5	2	4	L	500
48	5	2	4	L	1000
49	5	2	4	M	100
50	5	2	4	M	500
51	5	2	4	M	1000
52	5	2	4	W	100
53	5	2	4	W	500
54	5	2	4	W	1000
55	5	4	2	L	100
56	5	4	2	L	500
57	5	4	2	L	1000
58	5	4	2	M	100
59	5	4	2	M	500
60	5	4	2	M	1000
61	5	4	2	W	100
62	5	4	2	W	500
63	5	4	2	W	1000
64	5	4	4	L	100
65	5	4	4	L	500
66	5	4	4	L	1000
67	5	4	4	M	100
68	5	4	4	M	500
69	5	4	4	M	1000
70	5	4	4	W	100
71	5	4	4	W	500
72	5	4	4	W	1000
73	6	2	2	L	100
74	6	2	2	L	500
75	6	2	2	L	1000
76	6	2	2	M	100
77	6	2	2	M	500
78	6	2	2	M	1000
79	6	2	2	W	100
80	6	2	2	W	500
81	6	2	2	W	1000

TABLE 1
(continued)

Cell No.	No. of Variables	No. of Classes	No. of Groups	Discrimination* Level	Sample Size
82	6	2	4	L	100
83	6	2	4	L	500
84	6	2	4	L	1000
85	6	2	4	M	100
86	6	2	4	M	500
87	6	2	4	M	1000
88	6	2	4	W	100
89	6	2	4	W	500
90	6	2	4	W	1000
91	6	4	2	L	100
92	6	4	2	L	500
93	6	4	2	L	1000
94	6	4	2	M	100
95	6	4	2	M	500
96	6	4	2	M	1000
97	6	4	2	W	100
98	6	4	2	W	500
99	6	4	2	W	1000
100	6	4	4	L	100
101	6	4	4	L	500
102	6	4	4	L	1000
103	6	4	4	M	100
104	6	4	4	M	500
105	6	4	4	M	1000
106	6	4	4	W	100
107	6	4	4	W	500
108	6	4	4	W	1000

*L = Large Discrimination
M = Medium Discrimination
W = Weak Discrimination

TABLE 2

INDEPENDENT VARIABLES IN ANALYSIS OF BIAS

1. The Number of Variables (NVAR)
2. The Number of Classes (NCLASS)
3. The Number of Groups (NGROUP)
4. The Discriminatory Power of the Variable (DISC)
5. The Sample Size (NCASE)

TABLE 3

ANOVA FOR ITER

Source*	Sum of Squares	Degrees of Freedom	Mean Square	F	Tail Prob.
Mean	44535227.83135	1	44535227.83135	14805.29	0.0000
G	1547471.17485	2	773735.58742	257.22	0.0000
H	28768481.78002	1	28768481.78002	9563.79	0.0000
I	1285.42211	1	1285.42211	.43	.5133
J	4073182.73969	2	2036591.36985	677.04	0.0000
K	42163759.12220	2	21081879.56110	7008.46	0.0000
GH	1389742.78914	2	694871.39457	231.00	0.0000
GI	23534.35045	2	11767.17523	3.91	0.0200
HI	848.52353	1	848.52353	.28	0.5954
GJ	457166.73119	4	114291.68220	38.00	0.0000
HJ	4975965.01642	2	2487982.50821	827.10	0.0000
IJ	38884.56298	2	19442.28149	6.46	0.0016
GK	2059733.17030	4	514933.29258	171.18	0.0000
HK	28356932.28173	2	14178466.14087	4713.49	0.0000
IK	1678.30359	2	839.15180	.28	0.7566
JK	9853533.70591	4	2463383.42648	818.93	0.0000
GHI	24091.75861	2	12045.87931	4.00	0.0183
GHJ	479995.59371	4	119998.89843	39.89	0.0000
GHK	1852062.17874	4	463015.54469	153.92	0.0000
GIJ	34237.55514	4	8559.38878	2.85	0.0226
GIK	33572.00760	4	8393.00190	2.79	0.0249
HIJ	41152.94226	2	20576.47113	6.84	0.0011
HIK	1104.12521	2	552.06260	.18	0.8323
GJK	813430.25351	8	101678.78169	33.80	0.0000
HJK	11460064.52429	4	2865016.13107	952.45	0.0000
IJK	50897.04686	4	12724.26171	4.23	0.0020
GHIJ	34052.85908	4	8513.21477	2.83	0.0232
GHIK	34043.06728	4	8510.76682	2.83	0.0233
GHJK	8624 8.09836	8	107804.76230	35.84	0.0000
GIJK	50599.71536	8	6324.96442	2.10	0.0321
HIJK	55049.08429	4	13762.27107	4.58	0.0011
GHIJK	50831.59420	8	6353.94927	2.11	0.0313
Error	29824939.45319	9915	3008.06248		

*G=NVAR
H=NCLASS
I=NGROUP
J=NCASE
K=DISC

TABLE 4

ANOVA FOR CONBIAS

Source*	Sum of Squares	Degrees of Freedom	Mean Square	F	Tail Prob.
Mean	101.00894	1	101.00394	33147.14	0.0000
G	.20606	2	.10253	33.65	0.0000
H	22.39927	1	22.39927	7350.92	0.0000
I	.00964	1	.00964	3.16	0.0753
J	14.95401	2	7.47701	2453.78	0.0000
K	21.40436	2	10.70218	3512.21	0.0000
GH	.44777	2	.22389	73.47	0.0000
GI	.00586	2	.00293	.96	0.3825
HI	.01001	1	.01001	3.28	0.0700
GJ	.30757	4	.07689	25.23	0.0000
HJ	.14925	2	.07462	24.49	0.0000
IJ	.00146	2	.00073	.24	0.7875
GK	.52722	4	.13181	43.26	0.0000
HK	15.29812	2	7.64906	2510.24	0.0000
IK	.01762	2	.00881	2.89	0.0555
JK	.20342	4	.05086	16.69	0.0000
GHI	.00478	2	.00239	.78	0.4563
GHJ	.25501	4	.06375	20.92	0.0000
GHK	.63339	4	.15835	51.97	0.0000
GIJ	.01984	4	.00496	1.63	0.1643
GIK	.00600	4	.00150	.49	0.7417
HIJ	.00127	2	.00063	.21	0.8124
HIK	.01833	2	.00917	3.01	0.0494
GJK	.47943	8	.05993	19.67	0.0000
HJK	.89746	4	.22437	73.63	0.0000
IJK	.00586	4	.00145	.48	0.7502
GHIJ	.02230	4	.00558	1.83	0.1200
GHIK	.00473	4	.00118	.39	0.8175
GHJK	.43015	8	.05377	17.65	0.0000
GIJK	.02958	8	.00370	1.21	0.2863
HIJK	.00537	4	.00134	.44	0.7791
GHIJK	.03310	8	.00414	1.36	0.3098
Error	30.21238	9915	.00305		

*G=NVAR
H=NCLASS
I=NGROUP
J=NCASE
K=DISC

TABLE 5

ANOVA FOR VARBIAS

Source*	Sum of Squares	Degrees of Freedom	Mean Square	F	Tail Prob.
Mean	37.55725	1	37.55725	39017.73	0.0000
G	.01702	2	.00851	8.84	0.0001
H	6.23191	1	6.23191	6474.25	0.0000
I	.00161	1	.00161	1.68	0.1953
J	10.11093	2	5.05547	5252.06	0.0000
K	4.13598	2	2.06799	2148.41	0.0000
GH	.01939	2	.00970	10.07	0.0000
GI	.00242	2	.00121	1.25	0.2852
HI	.00169	1	.00169	1.75	0.1858
GJ	.02685	4	.00671	6.97	0.0000
HJ	.71491	2	.35746	371.36	0.0000
IJ	.00071	2	.00035	.37	0.6925
GK	.03065	4	.00764	7.94	0.0000
HK	2.58313	2	1.29156	1341.79	0.0000
IK	.00424	2	.00212	2.20	0.1108
JK	.28145	4	.07036	73.10	0.0000
GHI	.00247	2	.00124	1.28	0.2771
GHJ	.03205	4	.00801	8.32	0.0000
GHK	.03211	4	.00803	8.34	0.0000
GIJ	.00571	4	.00143	1.48	0.2045
GIK	.00255	4	.00064	.66	0.6188
HIJ	.00061	2	.00030	.32	0.7294
HIK	.00463	2	.00231	2.40	0.0904
GJK	.02954	8	.00369	3.84	0.0002
HJK	.13400	4	.03350	34.80	0.0000
IJK	.00078	4	.00020	.20	0.9366
GHIJ	.00519	4	.00130	1.35	0.2496
GHIK	.00257	4	.00064	.67	0.6142
GHJK	.02956	8	.00370	3.84	0.0002
GIJK	.00848	8	.00106	1.10	0.3589
HIJK	.00103	4	.00026	.27	0.8994
GHIJK	.00777	8	.00097	1.01	0.4268
Error	9.54387	9915	.00096		

*G=NVAR
H=NCLASS
I=NGROUP
J=NCASE
K=DISC

TABLE 6

ANOVA FOR CATBIAS

Source*	Sum of Squares	Degrees of Freedom	Mean Square	F	Tail Prob.
Mean	9.96646G	1	9.96646	39039.35	0.0000
G	.02812	2	.01406	55.06	0.0000
H	1.48303	1	1.48303	5809.13	0.0000
I	.00000	1	.00000	.01	0.9353
J	2.74748	2	1.37374	5381.03	0.0000
K	1.74328	2	.87164	3414.27	0.0000
GH	.02663	2	.01331	52.16	0.0000
GI	.00006	2	.00003	.11	0.8938
HI	.00000	1	.00000	.01	0.9384
GJ	.01671	4	.00418	16.37	0.0000
HJ	.19408	2	.09702	380.02	0.0000
IJ	.00034	2	.00017	.66	0.5186
GK	.03665	4	.00916	35.89	0.0000
HK	.84843	2	.42421	1661.67	0.0000
IK	.00005	2	.00003	.10	0.9044
JK	.25104	4	.06276	245.83	0.0000
GHI	.00006	2	.00003	.12	0.8879
GHJ	.01590	4	.00397	15.57	0.0000
GHK	.03788	4	.00947	37.10	0.0000
GIJ	.00083	4	.00021	.82	0.5150
GIK	.00013	4	.00003	.13	0.9710
HIJ	.00032	2	.00016	.63	0.5348
HIK	.00005	2	.00002	.10	0.9078
GJK	.02936	8	.00367	14.38	0.0000
HJK	.04105	4	.01026	40.20	0.0000
IJK	.00115	4	.00029	1.13	0.3399
GHIJ	.00082	4	.00020	.80	0.5255
GHIK	.00013	4	.00003	.12	0.9737
GHJK	.02840	8	.00355	13.91	0.0000
GIJK	.00108	8	.00013	.53	0.8363
HIJK	.00110	4	.00028	1.08	0.3635
GHIJK	.00103	8	.00013	.51	0.8524
Error	2.53123	9915	.00026		

*G=NVAR
H=NCLASS
I=NGROUP
J=NCASE
K=DISC

TABLE 7

ANOVA FOR GBIAS

Source*	Sum of Squares	Degrees of Freedom	Mean Square	F	Tail Prob.
Mean	11.36148	1	11.36148	24210.62	0.0000
G	.03999	2	.01999	42.61	0.0000
H	.64274	1	.64274	1369.65	0.0000
I	.24121	1	.24121	514.00	0.0000
J	2.15318	2	1.07659	2294.15	0.0000
K	1.95062	2	.97531	2078.32	0.0000
GH	.01361	2	.00680	14.50	0.0000
GI	.00011	2	.00006	.12	0.8888
HI	.00787	1	.00787	16.77	0.0000
GJ	.00450	4	.00112	2.40	0.0481
HJ	.04981	2	.02491	53.07	0.0000
IJ	.11146	2	.05573	118.76	0.0000
GK	.03614	4	.00904	19.25	0.0000
HK	1.00499	2	.50249	1070.78	0.0000
IK	.00791	2	.00395	8.43	0.0002
JK	.03873	4	.00968	20.63	0.0000
GHI	.00084	2	.00042	.89	0.4107
GHJ	.01101	4	.00275	5.86	0.0001
GHK	.01798	4	.00450	9.58	0.0000
GIJ	.00522	4	.00130	2.78	0.0253
GIK	.00057	4	.00014	.30	0.8753
HIJ	.00174	2	.00087	1.86	0.1564
HIK	.00487	2	.00244	5.19	0.0056
GJK	.00626	8	.00078	1.67	0.1011
HJK	.06534	4	.01634	34.81	0.0000
IJK	.00151	4	.00038	.81	0.5216
GHIJ	.00151	4	.00038	.81	0.5216
GHIK	.00081	4	.00020	.43	0.7860
GHJK	.01849	8	.00231	4.93	0.0000
GIJK	.00479	8	.00060	1.28	0.2508
HIJK	.00130	4	.00033	.69	0.5967
GHIJK	.00461	8	.00058	1.23	0.2785
Error	4.65288	9915	.00047		

*G=NVAR
H=NCLASS
I=NGROUP
J=NCASE
K=DISC

TABLE 8

ANOVA FOR ZKBIAS

Source*	Sum of Squares	Degrees of Freedom	Mean Square	F	Tail Prob.
Mean	4553.33104	1	4553.33104	3513549.71	0.0000
G	.00055	2	.00028	.21	0.8083
H	.08606	1	.08606	66.41	0.0000
I	.00000	1	.00000	.00	0.9887
J	.09134	2	.04567	35.24	0.0000
K	.71391	2	.35695	275.44	0.0000
GH	.02856	2	.01428	11.02	0.0000
GI	.00454	2	.00227	1.75	0.1739
HI	.00376	1	.00376	2.90	0.0887
GJ	.01468	4	.00367	2.83	0.0231
HJ	.17485	2	.08743	67.46	0.0000
IJ	.01559	2	.00780	6.02	0.0024
GK	.03282	4	.00820	6.33	0.0000
HK	.70632	2	.35316	272.51	0.0000
IK	.00620	2	.00310	2.39	0.0913
JK	.23740	4	.05935	45.80	0.0000
GHI	.00282	2	.00141	1.09	0.3373
GHJ	.00538	4	.00134	1.04	0.3861
GHK	.03486	4	.00871	6.72	0.0000
GIJ	.00649	4	.00162	1.25	0.2864
GIK	.00352	4	.00088	.68	0.6064
HIJ	.00525	2	.00263	2.03	0.1320
HIK	.00389	2	.00195	1.50	0.2229
GJK	.00851	8	.00106	.82	0.5843
HJK	.35410	4	.08852	68.31	0.0000
IJK	.00303	4	.00076	.58	0.6744
GHIJ	.00465	4	.00116	.90	0.4647
GHIK	.00790	4	.00198	1.52	0.1920
GHJK	.01812	8	.00226	1.75	0.0824
GIJK	.00668	8	.00083	.64	0.7412
HIJK	.00895	4	.00224	1.73	0.1409
GHIJK	.00565	8	.00071	.54	0.8235
Error	12.84919	9915	.00130		

*G=NVAR
H=NCLASS
I=NGROUP
J=NCASE
K=DISC

TABLE 9

ANOVA FOR TOTBIAS

Source*	Sum of Squares	Degrees of Freedom	Mean Square	F	Tail Prob.
Mean	56.09501	1	56.09501	78353.40	0.0000
G	.06346	2	.03173	44.32	0.0000
H	10.53650	1	10.53650	14717.36	0.0000
I	.31199	1	.31199	435.79	0.0000
J	12.51247	2	6.25624	8738.70	0.0000
K	5.45249	2	2.72624	3808.01	0.0000
GH	.03063	2	.01531	21.39	0.0000
GI	.01903	2	.00951	13.29	0.0000
HI	.08806	1	.08806	123.00	0.0000
GJ	.03445	4	.00861	12.03	0.0000
HJ	.71906	2	.35953	502.19	0.0000
IJ	.07724	2	.03862	53.94	0.0000
GK	.06544	4	.01636	22.85	0.0000
HK	4.81872	2	2.40936	3365.39	0.0000
IK	.01751	2	.00875	12.23	0.0000
JK	.04839	4	.01210	16.90	0.0000
GHI	.00709	2	.00354	4.95	0.0071
GHJ	.03064	4	.00766	10.70	0.0000
GHK	.02993	4	.00748	10.45	0.0000
GIJ	.01052	4	.00263	3.67	0.0054
GIK	.00411	4	.00103	1.44	0.2189
HIJ	.01839	2	.00919	12.84	0.0000
HIK	.00826	2	.00413	5.77	0.0031
GJK	.06293	8	.00787	10.99	0.0000
HJK	.06129	4	.01532	21.40	0.0000
IJK	.00650	4	.00162	2.27	0.0593
GHIJ	.00737	4	.00184	2.58	0.0357
GHIK	.00294	4	.00073	1.03	0.3921
GHJK	.05200	8	.00650	9.08	0.0000
GIJK	.01114	8	.00139	1.94	0.0493
HIJK	.00773	4	.00193	2.70	0.0290
GHIJK	.01177	8	.00147	2.06	0.0366
Error	7.09838	9915	.00072		

*G=NVAR
H=NCLASS
I=NGROUP
J=NCASE
K=DISC

TABLE 10

BOOTSTRAP RESULTS FOR CELL 89
2 CLASS/4 GROUP/6 VARIABLE/500 SAMPLE SIZE/WEAK DISCRIMINATION

	True parameter	Bootstrap estimate	Standard error	Minimum value	Maximum value	Bootstrap bias	Worst-case bias
CLASS 1							
Group 1	.5000	.4983	.0054	.3686	.6474	.0017	.1474
2	.5000	.5050	.0050	.3917	.6386	.0050	.1386
3	.5000	.4972	.0049	.3583	.6374	.0028	.1417
4	.5000	.4920	.0047	.3797	.6190	.0080	.1283
VAR 1							
Mean	5.0000	5.0164	.0084	4.7679	5.1990	.0164	.2321
S.D.	1.0000	.9968	.0055	.8477	1.1305	.0032	.1523
VAR 2							
Mean	5.0000	4.9984	.0081	4.7050	5.1648	.0016	.295
S.D.	1.0000	.9876	.0057	.8775	1.1104	.0124	.1225
VAR 3							
Level 1	.6000	.5972	.0033	.5105	.6778	.0028	.0895
VAR 4							
Level 1	.6000	.6007	.0030	.5275	.6901	.0007	.0901
VAR 5							
Level 1	.6000	.6009	.0036	.4540	.6962	.0009	.146
VAR 6							
Level 1	.6000	.5942	.0029	.5285	.6825	.0058	.0825

TABLE 10
(continued)

	True parameter	Bootstrap estimate	Standard error	Minimum value	Maximum value	Bootstrap bias	Worst-case bias
CLASS 1							
Group 1	.5000	.5017	.0054	.3526	.6314	.0017	.1474
2	.5000	.4950	.0050	.3614	.6083	.0050	.1386
3	.5000	.5028	.0049	.3626	.6417	.0028	.1417
4	.5000	.5080	.0047	.3810	.6203	.0080	.1203
VAR 1							
Mean	3.0000	3.0005	.0081	2.7583	3.2069	.0005	.2417
S.D.	1.0000	1.0058	.0055	.8693	1.1475	.0058	.1475
VAR 2							
Mean	3.0000	3.0024	.0080	2.7891	3.2107	.0024	.2109
S.D.	1.0000	1.0014	.0062	.8628	1.1563	.0014	.1563
VAR 3							
Level 1	.4000	.3961	.0034	.2994	.4876	.0039	.1006
VAR 4							
Level 1	.4000	.4024	.0032	.3083	.4776	.0024	.0917
VAR 5							
Level 1	.4000	.3946	.0037	.3041	.4686	.0054	.0959
VAR 6							
Level 1	.4000	.4054	.0034	.3280	.5060	.0054	.1060

TABLE 11

NUMBER OF REPETITIONS WITHIN CELLS OF THE DESIGN

Cell No.	Repetitions	Cell No.	Repetitions	Cell No.	Repetitions
1	100	37	100	73	100
2	100	38	100	74	100
3	100	39	100	75	100
4	100	40	100	76	100
5	100	41	100	77	100
6	100	42	100	78	100
7	100	43	100	79	100
8	100	44	100	80	100
9	100	45	100	81	100
10	100	46	100	82	100
11	100	47	100	83	100
12	100	48	100	84	100
13	100	49	100	85	100
14	100	50	100	86	100
15	100	51	100	87	100
16	100	52	100	88	100
17	100	53	100	89	100
18	100	54	100	90	100
19	100	55	100	91	100
20	100	56	100	92	100
21	100	57	100	93	100
22	100	58	100	94	100
23	100	59	100	95	100
24	100	60	100	96	100
25	97	61	65	97	100
26	41	62	42	98	93
27	9	63	11	99	45
28	100	64	100	100	100
29	100	65	100	101	100
30	100	66	100	102	100
31	100	67	100	103	99
32	100	68	100	104	100
33	100	69	100	105	100
34	96	70	99	106	99
35	52	71	22	107	97
36	14	72	1	108	41

CHAPTER V

APPLICATIONS

This chapter suggests two applications of the latent discriminant model. They are: (1) perceptual mapping and (2) constrained clustering. The chapter is organized as follows. First the empirical data are described. Next, each application is discussed in turn. For each application, an operationalization will be described via the latent discriminant model. After that an empirical example will be presented. The chapter ends by summarizing and suggesting other areas of application.

The Data

The data consist of perceptual, psychographic, and demographic data from 985 respondents. Each respondent was asked for responses on attributes of bath soap for their most used brand, second most used brand (if any), third most used brand (if any), and a "test" brand. All respondents were exposed to the test brand. In addition, demographic and psychographic information were collected for each respondent. Table 12 gives a complete description of the scales and items.

Perceptual Mapping

The Model

Perceptual mapping has previously been operationalized in the Marketing literature through a variety of analytic methods and models.

The methods commonly used are Factor Analysis (e.g., Hauser and Urban 1977, Hauser and Gaskin 1984), Multiple Discriminant Analysis (Johnson 1971, Pessemier 1975), Principal Components Analysis (Holbrook and Huber 1979), and Multidimensional Scaling (Green and Carmone 1970, Green and Rao 1972). In a prototypical perceptual mapping application, individuals are scaled along some underlying continuum with respect to a set of brands. The data required for this exercise can be pictured as a three-way array. If we have N individuals assessing J objects on P attributes, our three way matrix $\underset{\sim}{A}$ has elements defined as

(58) a_{ijk}; $[i=1,2,\ldots,N;\ j=1,2,\ldots,J;\ k=1,2,\ldots,P]$

The data cube defined in (58) has three modes, three indices by which the data can be classified. The Latent Discriminant Method accepts as input the raw data matrix $\underset{\sim}{A}$. Each attribute may be measured on a nominal, ordinal or interval scale. Note that one can mix the three types of scales in a single Latent Discrimination application. For the purpose of explicating the Latent Discriminant Analysis model, it would be convenient to picture the data in a slightly different manner. Let the J brands denote the groups. Assuming for the present that each of the N individuals in our sample evaluates each of the J brands, we have NxJ evaluations. Each of the NxJ evaluations is represented by a vector y corresponding to the attributes in question. Let G_j denote the group of interest (G_j stands for the j^{th} brand). We can represent the data as a matrix of $[G|Y]$, where G is an (NxJ)x1 column vector indicating the evaluations of J brands by each of the N individuals and Y is an (NxJ)xP matrix of measurements for each of the N individuals.

The Model will be developed by means of a simple example. Assume that we have asked N individuals to evaluate two brands (G_1 and G_2) on four attributes (A,B,C,D). We have chosen A, B, C, and D such that A and B are the manifest indicators of one perceptual dimension of the brands in question (say D_1) and C and D reflect another underlying dimension (say D_2). As mentioned before, each dimension is a continuum. Assume that the continuum may be approximated by two classes: A "High" class and a "Low" class. Therefore, the perceptual space for our two dimensional situation (D_1D_2) may be fully described by four classes: (1) a (High, High) class, (2) A (High, Low) class, (3) a (Low, High) class, and (4) a (Low, Low) class. We will denote these four classes as $t_1,\ldots,t_4$ respectively. t_1 and t_2 represent the classes that are High on D_1. To achieve this we must constrain $F(A|t_1) = F(A|t_2)$, and the same for B. Similarly as t_3 and t_4 are "Low" classes on D_1 and therefore we must restrict $F(A|t_3) = F(A|t_4)$, and the same for B. Similar restrictions will be applied to C and D with respect to D_2. Let Θ_{AT} be the conditional density of a given latent class t. We have:

(59) $P(y|G_j) = \sum_{t=1}^{4} F(A|\Theta_{AT})F(B|\Theta_{BT})F(C|\Theta_{CT})F(D|\Theta_{DT})\ P(t|G_j)$

subject to:

$\Theta_{A1} = \Theta_{A2},\quad \Theta_{A3} = \Theta_{A4}$

$\Theta_{B1} = \Theta_{B2},\quad \Theta_{B3} = \Theta_{B4}$

$\Theta_{C1} = \Theta_{C3},\quad \Theta_{C2} = \Theta_{C4}$

$\Theta_{D1} = \Theta_{D3},\quad \Theta_{D2} = \Theta_{D4}$

We now give distinct psychological meaning to the parameters. We have assumed that the perceptual space can be adequately described by T latent classes. The parameters $F(y_i|\Theta_{y_i t})$ now represent the marginal density of attribute y_i in latent class t. For any given value of y_i this can be loosely interpreted as the probability of observing that level of the attribute given the t^{th} latent class.

The parameters $P(t|G_j)$ is the probability of observing the t^{th} latent class given the brand in question is brand j. Depending on the attributes perceived by individuals in each of the brands, we would expect this probability to be different across classes. The perceptual mapping problem in this case is simply a mapping of each individual from $[G_j|Y]$ to $[\underset{\sim}{\Theta},t|G_j]$. The solution in terms of the model described above gives us a perceptual space for the group as a whole.

Figure 5.1 helps clarify the latent perceptual mapping strategy. We divide our two dimensional latent perceptual space into four quadrants. Each quadrant corresponds to a latent class. The relative size of each quadrant is determined by the latent class proportions. For each latent class (quadrant) we obtain the mean attribute levels, and more interestingly, the probability of observing the latent class (quadrant) given a brand. This preserves the "group" effect that two mode methods do not preserve.

An advantage of latent perceptual mapping is that it preserves individual differences. We can scale individuals along each dimension with respect to the brands perceived by the individual. A method of scaling individuals is now described. We are given a vector of perceptions with respect to an individual (y,G_j) and the latent perceptual solution from (2). Referring back to Figure 5.1, we wish to scale the individual along D_1. The probability of the individual being "High" ($\ell 1$) on dimension D_1 with respect to brand j = $[P(t_1|\mathcal{Y},G_j)+P(t_2|\mathcal{Y},G_j)]$ = $P(D_{\ell 1}|\mathcal{Y},G_j)$. The probability of the individual being "low" ($\ell 2$) on dimension D1 with respect to brand j = $[P(t_3|\mathcal{Y},G_j)+P(t_4|\mathcal{Y},G_j)]$ = $P(D_{\ell 2}|\mathcal{Y},G_j)$. The $P(D_{\ell i}|\mathcal{Y},G_j)$ give us a value for the level of the dimension the individual probably came from without using the order implicit in the model. Let $Z_{\ell i}$ represent the proportion of responses "below" level "ℓ" of dimension i. For the "high" level, this is simply equal to the sum of the latent class proportions $(P(X_t))$ for all lower levels of the dimension, with the convention that the value of $Z_{\ell i}$ for the highest level = 1 and the value for the lowest level = 0. Then, adapting from Lazarsfeld and Henry's discussion on ordered latent classes (1968) we construct a scale $S_i(y,G_j)$ with respect to dimension D_i having L levels as follows

$$(60)\quad S_i(\mathcal{Y},G_j) = \sum_{\ell=2}^{L} Z_{\ell i} P(D_{\ell i}|y,G_j) + \sum_{\ell=1}^{L} P(D_{\ell i})P(D_{\ell i}|\mathcal{Y},G_j)/2 \quad .$$

Computing the above scale value for each dimension gives us the perceptual coordinates for a given individual for a given brand.

An Illustration

We use the data described in Table 12 to define our perceptual space. For illustrative purposes, we restrict the discussion to a two dimensional perceptual space. Dimension 1, a "clean" dimension, is defined by two manifest indications: "Leaves me feeling refreshed" (called "Var A"), and "Leaves me feeling clean" (called Var B). Dimension 2, a "deodorant protection" dimension, is defined by a single indicator, "is a soap that provides deodorant protection" (called Var C). Responses were on a 10 point scale, with a 10 corresponding to "agree strongly" and a 1 corresponding to "disagree strongly." Perceptions were measured on the brand used most often, a "test" brand, the second most often used brand (if any) and the most used brand (if any). All respondents volunteered perceptions on the test brand, however perceptual data for the other brands varied across respondents. Nine brands were included in the analysis.

Estimating the Model

To estimate the model, constraints mentioned in (5.2) must be in place. Let $f(A)_j$ denote the conditional estimate of indicator A in latent class j. We use a four class model to estimate the perceptual space. The constraints are:

$$f(A)_1 = f(A)_2,\ f(A)_3 = f(A)_4$$
$$f(B)_1 = f(B)_2,\ f(B)_3 = f(B)_4$$
$$f(C)_1 = f(C)_3,\ f(C)_2 = f(C)_4$$

The estimated model results are shown in Table 13. The mixture proportions give us the probability of observing a latent class (quadrant) given the group (brand) in question. Brand 9 was the "test" brand. Note that the probability of observing Latent Class 4 given brand 9 was .5213. In other words, the probability of perceiving the test brand as Low on "clean" and Low on "deodorant" was .5213. Latent class 1 corresponds to the "High,High" quadrant in Figure 5.1; latent class 2 corresponds to the "High,Low" quadrant in Figure 5.1; latent class 3 corresponds to the "Low,High" quadrant in Figure 5.1; and latent class 4 corresponds to the "Low,Low" quadrant in Figure 5.1.

Scaling individuals. We now construct an individual's perceptual map. Perceptions of an individual on three brands are given in Table 14. We will derive the individual coordinates for the brand used most often (Brand 5 for this individual) in some detail, and then display the perceptual map.

We first obtain the coordinate for the cleanliness dimension. Referring to Table 13, we see that the probability of the individual being "High" on the cleanliness dimension is equal to the sum of the posterior probabilities of observing latent classes 1 and 2 given the individual's perceptual profile for Brand 5, i.e., .65504+.31882=.97386. The probability of the individual being "Low" on the cleanliness dimension is equal to the sum of the posterior probability of observing latent classes 3 and 4 given the individual's perceptual profile for Brand 5, i.e., 0+.026132=.026132. Using the convention for Z_h mentioned previously, we set Z_h=1. The scale value for the "clean" dimension on Brand 5 this is:

$$(1)(.97386)+[(.5185)(.97386)]/2+[(.4816)(.026132)]/2 = 1.23263$$

Now, for the deodorant dimension, the probability of the individual being "High" on this dimension given the individual's perceptual profile for Brand 5 is equal to the sum of the posterior probabilities of observing latent classes 1 and 3 given the individual's perceptual profile for Brand 5, i.e., .65504+0=.65504. The probability of the individual being "Low" on the cleanliness dimension is equal to the sum of the posterior probabilities of observing latent classes 2 and 4, i.e., .318882+.02613=.34495. By convention Z_h=1. The scale value for the "deodorant" dimension on Brand 5 is:

$$(1)(.65504)+[(.4853)(.65504)]/2+[(.5148)(.344954)]/2=.90278.$$

The coordinates for the "cleanliness" and the "deodorant" dimensions for Brand 5 are therefore (1.23263,.90278). Similarly, given the individual's perceptual profile, we calculate the coordinates for the "test" brand (Brand 9) as (12408,.2923) and for the second most used brand (Brand 8) as (.2408,.2574). Figure 5.2 displays the perceptual map.

Cluster Analysis

Cluster Analysis aims to classify objects into subgroups such that the subgroups show similarity with respect to some attributes of interest. A typical clustering application has as input data p measurements on each of n objects. The nxp raw data matrix is converted into an (nxn) similarity or distance matrix. A given clustering algorithm then allocates each of the n objects to one of k (k>1) groups. A typical clustering algorithm maximizes some distance measure across groups and minimizes the within group distance (the distance chosen is typically the Euclidean metric). The within group profiles are then examined in terms of their average values on other characteristics of interest

(Dillon and Goldstein 1984). As opposed to the case of discriminant analysis, classification is done purely by the given similarity (distance) rule. There is no "a priori" knowledge of the group membership of each profile. The similarity measures used are of two types: for data having metric properties, a distance-type measure can be used. With data having qualitative components, a matching-type measure is used. Let Y_i denote the measurements collected on the i^{th} object or individual with respect to p variables. Most distance measures are special cases of the Minkowski metric:

$$(61)\quad d_{ij} = \{\textstyle\sum_{k=1}^{p} |y_{ik}-y_{jk}|^{r}\}^{(1/r)}$$

where d_{ij} is the distance between object i and j. With r=2 we have the Euclidean distance formula. Matching-type measures are appropriate when the data are nominally scaled. They are also known as association coefficients. Association coefficients generally take on values between 0 and 1 and are based on the reasoning that two individuals should be judged similar to the extent that they share common attributes (Dillon and Goldstein 1984).

Constrained Clustering

Very often, the analyst may possess some information that imposes constraints on the classification. A simple example in marketing would be the assignation of client to salesmen, with the constraint that the clients are geographically contiguous and each salesman gets roughly the same dollar value in terms of sales potential. DeSarbo and Mahajan

(1984) have proposed an algorithm that allows one to create clusters based on external constraints. Their algorithm is a distance based procedure which uses a penalty function method to impose the constraints. A method of constrained clustering using the Latent Discriminant Analysis is proposed. Various types of constraints which may be operationalized are shown.

The Latent Clustering Model

The standard form of the Latent Discriminant Analysis is

$$P(y|G_k) = \sum_{t=1}^{T} P(y|X_t)P(X_T|G_K)$$

In the clustering case we have no a priori knowledge of group membership. The model therefore reduces to:

$$P(y) = \sum_{t=1}^{T} P(y|X_t)P(X_t)$$

which is the standard Latent Structure Model described in Chapter II. In the above model:

Let t=1, ..., T clusters (latent classes)

i=1, ..., p variables

j=1, ..., n objects

$P(X_t)$ = the proportion of objects in cluster T

e_{jt} = degree of membership of object j in cluster T

For various types of cluster applications, some constraints which can be operationalized in the Latent Clustering framework are mentioned. Similar constraints have been suggested by Mahajan and Jain (1978) and DeSarbo and Mahajan (1984):

a) $e_{jt}(1-e_{jt}) = 0$ $j = 1, \ldots, N,$ $t=1, \ldots, T.$

$\sum_{t=1}^{T} e_{jt}=1$ $j=1, \ldots, N$

The above constraint provides for non-overlapping clusters, where objects can belong to one and only one cluster. This is easily operationalized in our framework by adapting the likelihood:

$$L = \sum_{j=1}^{N} \sum_{t=1}^{T} I_{jt} P(y|X_t) P(X_t)$$

Where I_{jt} is an indicator variable equal to 1 for that cluster to which the object belongs, and zero otherwise.

b) $0 \leq e_{jt} \leq 1$ $j=1, \ldots, N$ $t=1, \ldots T$

$\sum_{t=1}^{T} e_{jt}=1.$ $j=1, \ldots, N.$

These constraints allow for "fuzzy-set" clusters where objects can be fractional members of clusters. This is simply achieved by removing the indicator variable from the likelihood equation stated above. Objects will now be assigned to clusters in proportion to the probability of observing the object given the cluster

c) $e_{jt} + e_{kt} = 2.$

Here one wants objects j and k to belong to the same cluster t. This can simply be operationalized by creating an artificial variable z with two levels, 0 and 1. The variable z will be set equal to 1 for those objects desired in the same cluster. Then, for all clusters except the desired cluster t, set the conditional density of observing z=1 to zero.

d) $e_{mt} + e_{nt} \leq 1 \quad t=1, \ldots, T$

This is a generalization of c above. It forbids objects m and n to be in the same cluster. A variant of the same artificial variable ploy would be used.

e) $P(X_t) = P(X_t)$

This constraint restricts two cluster sizes to be equal. An equality constraint on the cluster proportions will achieve this constraint.

f) $P(X_t) \geq P(X_K)$

This inequality constraint allows one to have at least as many members in $P(X_t)$ as in $P(X_K)'$. The operationalization is as follows. "Create" two clusters $\{P(X_t)_1 P(X_t)_2\} \varepsilon P(X_t)$. Set $P(X_K)$ equal to $P(X_t)_1$. Set the conditional densities of $P(X_t)_1$ and $P(X_t)_2$ equal. The desired result is achieved.

(g) $p(y_i|X_t)=K$.

This constraint sets the estimate for attribute y_i given latent class t to some constant.

<u>An Illustration</u>

To illustrate the constrained clustering algorithm we give an example from the data described in Table 12. Specifically, we wish to test whether respondents differ with respect to the usage situation (i.e., do they use soap in the bath or soap in the shower) and employment status by their preference (or lack of it) for deodorant soaps. Items from the psychographic inventory used were the following. "A deodorant soap is the best way to get clean" (Var 1), "I use only a deodorant soap" (Var 2), "the only thing soaps should do is cleanse you--it is not necessary to deodorize or moisturize" (Var 3), "Deodorant soap gives you a confident feeling when you're with other people" (Var 4). All these items were measured on a 6 point scale. The usage situation variable (Var 5) has three nominal categories: (1) bath, (2) shower, (3) both equally, and the employment variable (Var 6) is measured at three nominal levels: (1) unemployed, (2) less than 35 hours, and (3) 35 hours or more. To test whether preference for deodorant soap does make a difference, we constrained the means for (Var 1, Var 2, Var 3, Var 4) at (4.5, 4.2, 2.0, 4.5) for latent class 1. The solution is shown in Table 15. Latent class 1 has a proportion of .5292. The mean profiles for the psychographic variables clearly define a class with a preference for deodorant soaps, while latent class 2 has a lower mean profile. The usage variable however shows almost no difference across classes, while the employment variable shows a slightly larger proportion of individuals are employed at 35 hours or more for the preferred deodorant class (.4317 vs. .3969) and less than 35 hours (.1928 vs. .1668). Usage and employment do not show differences across segments defined by preference for deodorant soap.

Summary

In this chapter we have illustrated two applications of the latent discriminant model, perceptual mapping and constrained clustering. The perceptual map obtained via LADI has a few distinct advantages over standard mapping techniques: (1) it is probabilistically scaled with a natural zero point, (2) the map is unique, (3) no rotation is permissible, (4) a single variable can define a dimension, and (5) dimensions may be defined for qualitative factors. Some issues need to be explored, such as pre-processing the data, and the number of levels necessary to obtain a good map. The second application, constrained clustering, has wide use in many areas of research. An interesting issue in clustering would be to estimate error rates for constrained clustering solutions.

FIGURE 5.1 - A LATENT PERCEPTUAL MAP

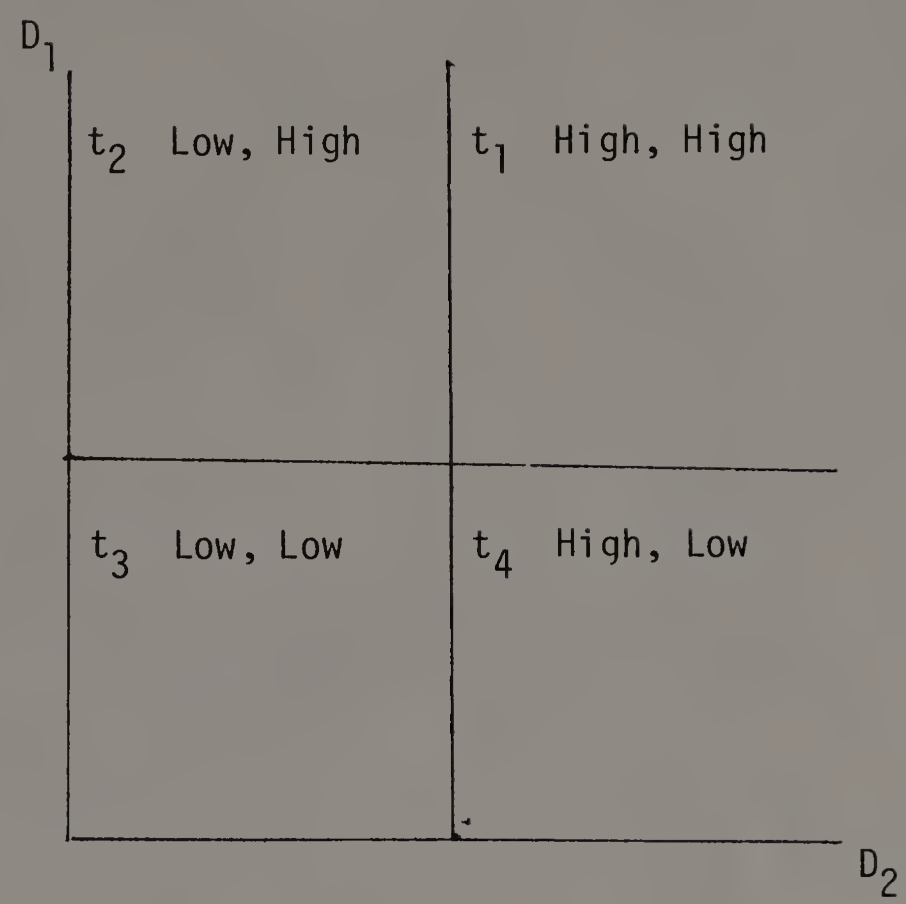

FIGURE 5.2 PERCEPTUAL MAP

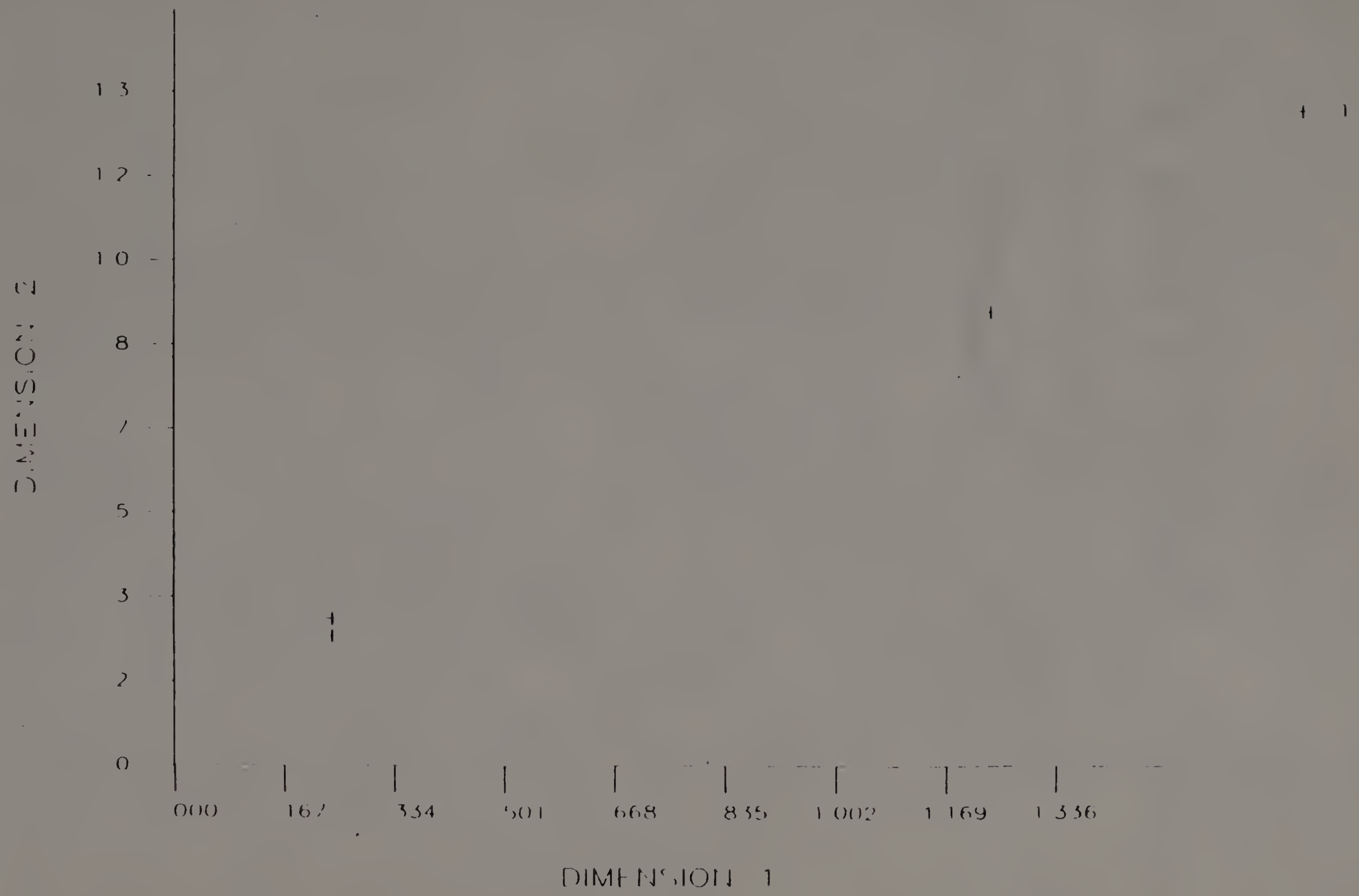

TABLE 12

DESCRIPTION OF SCALES AND ITEMS USED IN ANALYSIS

PSYCHOGRAPHIC ITEMS

Scale: 6 Strongly Agree
5 Somewhat Agree
4 Slightly Agree
3 Slightly Disagree
2 Somewhat Disagree
1 Strongly Disagree

I prefer cleansing creams to soaps for cleansing the face.

A deodorant soap is the best way to get really clean.

Dry skin isn't one of my problems.

I won't use any soap on my face.

Showers get you cleaner than baths.

I use only deodorant soap.

I prefer a soap with little or no added fragrance.

I often take a bath to relax.

I have a definite skin care program which I follow.

The only thing soaps should do is cleanse you--It is not necessary to deodorize or moisturize.

I often take a shower to relax me.

There are only certain soaps that I'll use on my face.

I doesn't matter much to me what brand of soap I use.

Deodorant soap gives you a confident feeling when you're with other people.

I prefer a soap whose fragrance stays on my skin.

TABLE 12
(continued)

DEMOGRAPHIC ITEMS

First would you say you usually take baths or do you usually take showers?

-1 Bath
-2 Shower
-3 Both equally

How many hours per week, if any, are you employed outside your home?

-1 None
-2 Less than 35 hours
-3 35 hours or more

PERCEPTUAL ITEMS

-1	1	Disagree Strongly
-2	2	
-3	3	
-4	4	
-5	5	
-6	6	
-7	7	
-8	8	
-9	9	
-0	10	Agree Strongly

Is a product I really like

Leaves me feeling refreshed

Has a pleasant fragrance

Is an all-family soap

Is a soap worth paying a little extra for

Is good for use in the shower

Is a soap for men to use

Rinses off easily

TABLE 12
(continued)

Is mild to the skin
Makes lots of rich and creamy lather
Is affordable for everyday use
Leaves me feeling clean
Is a quality product
Is a soap that provides deodorant protection
Has the right amount of fragrance
Is a luxury product
Is a soap for women to use
Is expensive
Is different from other soaps

TABLE 13

LATENT PERCEPTUAL MAPPING RESULT

Latent Class	1	2	3	4
Class Proportion $P(X_t)$	.3786	.1399	.1067	.3749
Mixture Proportions $(P(X_t/G_k))$				
Brand 1	.5890	.0822	.1770	.1518
2	.6074	.0206	.2766	.0955
3	.2227	.4643	.0000	.3130
4	.5469	.0000	.3246	.1285
5	.2394	.3256	.0000	.4350
6	.6888	.0000	.2627	.0485
7	.6474	.0700	.1044	.1783
8	.2604	.3414	.0048	.3934
9	.3186	.0712	.0889	.5213
Conditional Probabilities				
Var A				
Mean	9.15	9.15	5.01	5.01
	1.00	1.00	2.24	2.24
Var B				
Mean	9.80	9.80	6.26	6.26
	.40	.40	2.53	2.53
Var C				
Mean	9.20	4.79	9.20	4.79
	.91	2.31	.91	2.31

TABLE 14

PERCEPTUAL PROFILE OF THREE BRANDS
FOR A GIVEN INDIVIDUAL

	Brand No.	Leaves me feeling refreshed	Leaves me feeling clean	Provides deodorant protection
Brand used most often	5	9	9	8
Test brand	9	10	7	7
Brand used second most often	8	2	4	4

TABLE 15

LATENT CLUSTERING SOLUTION

	Cluster 1	Cluster 2
$P(X_t)$	.5292	.4708
Var 1		
Mean	4.5	2.05
S.D.	1.4	1.27
Var 2		
Mean	4.5	1.17
S.D.	1.45	.38
Var 3		
Mean	2.0	3.22
S.D.	1.56	1.85
Var 4		
Mean	4.5	2.23
S.D.	1.56	1.43
Var 5		
1	.2515	.2831
2	.6248	.6138
3	.1237	.1031
Var 6		
1	.3754	.4362
2	.1928	.1668
3	.4317	.3969

CHAPTER VI

SUMMARY

This chapter is divided into two sections. The first section restates the major results of this work, and the second outlines future areas of research.

Major Results

The problem considered is one of formulating a latent structure model that allows for overlapping groups and the ability to scale both categorical as well as continuous variables. Chapter II gives a methodological review that covers various models of the latent structure family that have been proposed.

In Chapter III we develop the Latent Discriminant model. Maximum Likelihood estimates are derived and an algorithm for solution is proposed. Some additional results are presented: a classification rule, check for identification, and a goodness-of-fit measure.

Chapter IV is devoted to a simulation that tests whether the algorithm suggested is capable of capturing a known structure. The simulation is carried out over 108 "true" parameter spaces and various performance measures are calculated to test the accuracy of the results. The results are informative in that they show that the algorithm works well under regular conditions, but is susceptible to small sample sizes and weak discrimination between classes. Chapter V presents two empirical applications of the Latent Discriminant Model, a Perceptual Mapping application and a constrained clustering application. The Perceptual Map is operationalized in the Latent Discriminant Framework, and an empirical result is shown. The Latent Discriminant Perceptual Map has certain advantages: (1) individual differences are preserved; (2) all respondents do not need to perceive an equal number of brands; (3) the dimensions may be qualitative; and (4) the dimensions are probabilistically scaled. Next, the constrained clustering model is operationalized in the Latent Discriminant Framework. Various possible constraints of interest are suggested and an empirical result is shown.

Future Work

This work has generated some interesting research opportunities and possibilities. For the model to be used in empirical research, work needs to be done to assure at least constrained global maximizers. Research by Hathaway (1983) on the constrained EM algorithm needs to be incorporated into the present framework. The feasibility of using the Latent Discriminant Model as a classification procedure needs to be explored. Finally, two Marketing Research problems are very amenable to being modeled in the Latent Discriminant Framework: (1) a "First Choice" model that allows for individual differences in both tastes and perceptions and (2) normative segmentation that incorporates both consumer level as well as resource constraints.

REFERENCES

Andersen, Erling B. (1982). "Latent Structure Analysis: A Survey," Scandinavian Journal of Statistics, 9, 1-12.

BMDP (1981). "BMDP Statistical Software," Dixon, W.J., ed. University of California Press: Berkeley.

Bonnett, Douglas G. and Peter M. Bentler (1980). "Goodness-of-Fit Procedures for the Evaluation and Selection of Log-Linear Models," Psychological Bulletin, 93, 149-166.

Carroll, J. Douglas and Jih-Jie Chang (1970). "Analysis of Individual Differences in Multidimensional Scaling Via an N-Way Generalization of 'Eckart-Young' Decomposition," Psychometrika, 35, 283-320.

Clogg, Clifford C. (1977). "Unrestricted and Restricted Maximum Likelihood Latent Structure Analysis: A Manual for Users," working paper 1977-09, Population Issues Research Office, Pennsylvania State University, University Park, PA 16802.

Clogg, Clifford C. (1979). "New Developments in Latent Structure Analysis," in "Factor Analysis and Measurement in Sociological Research: A Multidimensional Perspective," D.M. Jackson and E.F. Borgatta (eds.), Sage: Beverly Hills, CA.

Clogg, Clifford C. and Leo A. Goodman (1984). "Latent Structure Analysis of a Set of Multidimensional Contingency Tables." Journal of the American Statistical Association, 79, 762-771.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). "Maximum Likelihood Estimation From Incomplete Data Via the E-M Algorithm." Journal of the Royal Statistical Society B, 39, 1-22.

Desarbo, Wayne S. and Vijay Mahajan (1984). "Constrained Classification: The Use of A Priori Information in Cluster Analysis," Psychometrika, 49, 187-216.

Dillon, William R., and Matthew Goldstein (1984). Multivariate Analysis, Methods and Applications. New York: John Wiley.

Dillon, William R., Thomas J. Madden, and Narendra Mulani (1983). "Scaling Models for Categorical Variables: An Application of Latent Structure Models," Journal of Consumer Research, 10, 209-224.

Efron (1981). The Jackknife, the Bootstrap and Other Resampling Plans. Philadelphia: SIAM.

Eviritt, B.S. and D.J. Hand (1981). Finite Mixture Distributions. New York: Chapman and Hall.

Gibson, W.A. (1950). "Three Multivariate Models: Factor Analysis, Latent Structure Analysis, and Latent Profile Analysis," Psychometrika, 24, 229-252.

Goodman, Leo A. (1971). "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications," Technometrics, 13, 33-61.

Goodman, Leo A. (1972). "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables," American Sociological Review, 37, 28-46.

Goodman, Leo A. (1974a). "The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable. Part I: Modified Latent Structure Approach." American Journal of Sociology, 79, 1179-1259.

Goodman, Leo A. (1974b). "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." Biometrika, 61, 215-231.

Goodman, Leo A. and William H. Kruskal (1954). "Measures of Association for Cross Classifications." Journal of the American Statistical Association, 70, 755-768.

Green, Paul E. and Frank Carmone (1970). Multidimensional Scaling and Related Techniques in Marketing Analysis. Boston: Allyn and Bacon.

Green, Paul E., Frank J. Carmone, and David P. Wachpress (1976). "Consumer Segmentation via Latent Class Analysis," Journal of Consumer Research, 3, 170-174.

Green, Paul E. and Vithala R. Rao (1972). Applied Multidimensional Scaling. New York: Holt, Rinehart and Winston.

Grover, Rajiv and William Dillon (1984). "A Confirmatory Methodology for Market Structure Analysis," Marketing Science, under review.

Haberman, Shelby J. (1977). "Product Models for Frequency Table Involving Indirect Observation," Annals of Statistics, 5, 1124-1127.

Haberman, Shelby J. (1978). Analysis of Qualtative Data, Vol. I., New York: Academic Press.

Hathaway, Richard J. (1983). "Constrained Maximum-Likelihood Estimation for a Mixture of Multivariate Normal Distributions," Unpublished doctoral dissertation, Rice University.

Hauser, John R. and Steven P. Gaskin (1984). "Application of the 'Defender' Consumer Model," Marketing Science, 3, 327-351.

Hauser, John R. and Glen L. Urban (1977). "A Normative Methodology for Modelling Consumer Response to Innovation," Operations Research, 25 (July-August), 576-619.

Holbrook, Morris B. and Joel Huber (1979). "Separating Perceptual Dimensions from Affective Overtones: An Application to Consumer Aesthetics," Journal of Consumer Research, 5, 272-283.

Johnson, Richard M. (1971), "Market Segmentation: A Strategic Management Tool," Journal of Marketing Research, 8 (February), 13-18.

Jöreskog, Karl and Dag Sörbom (1981). "LISREL IV Analysis of Linear Structural Relationships by the Method of Maximum Likelihood," Chicago: Sci Ware.

Lazarsfeld, Paul (1959). "Latent Structure Analysis," in Psychology: A Study of a Science, S. Koch, ed., New York: McGraw Hill, 3, pp. 476-535.

Lazarsfeld, Paul F. and Neil W. Henry (1968). Latent Structure Analysis. Boston: Houghton-Mifflin.

Lord, Frederic and Melvin Novick (1966). Statistical Theories of Mental Test Scores," Reading, MA: Addison-Wesley.

Madden, Thomas J. and William R. Dillon (1982). "Causal Analysis and Latent Class Models: An Application to a Communications Hierarchy of Effects Model," Journal of Marketing Research, 19, 472-490.

Mahajan, Vijay and Arun K. Jain (1978). "An Approach to Normative Segmentation," Journal of Marketing Research, 15, 338-345.

McDonald, Roderick P. (1962). "A Note on the Derivation of the General Latent Class Model," Psychometrika, 27, 203-206.

McHugh, Richard B. (1956). "Efficient Estimation and Local Identification in Latent Class Analysis," Psychometrika, 23, 331-347.

Nicosia, Franco (1977). "Latent Structure Analysis," in Multivariate Methods for Market and Survey Research, J. Sheth, ed., Chicago: American Marketing Association.

Pessemier, Edgar A. (1975). "Market Structure Analysis of New Product and Market Opportunities," Journal of Contemporary Business (Spring), 35-67.

Poulsen, Carsten Stig (1982). "Latent Structure Analysis with Choice Modeling Applications," Unpublished doctoral dissertation, University of Pennsylvania.

Silk, Alvin J. and Glen L. Urban (1978). "Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," Journal of Marketing Research, 15, 171-191.

Sundberg, Rolf (1974). "Maximum Likelihood Theory for Incomplete Data From an Exponential Family," Scandinavian Journal of Statistics, 2, 1-49.

Wolfe, J.H. (1970). "Pattern Clustering by Multivariate Mixture Analysis," Multivariate Behavioral Research, 5, 329-350.